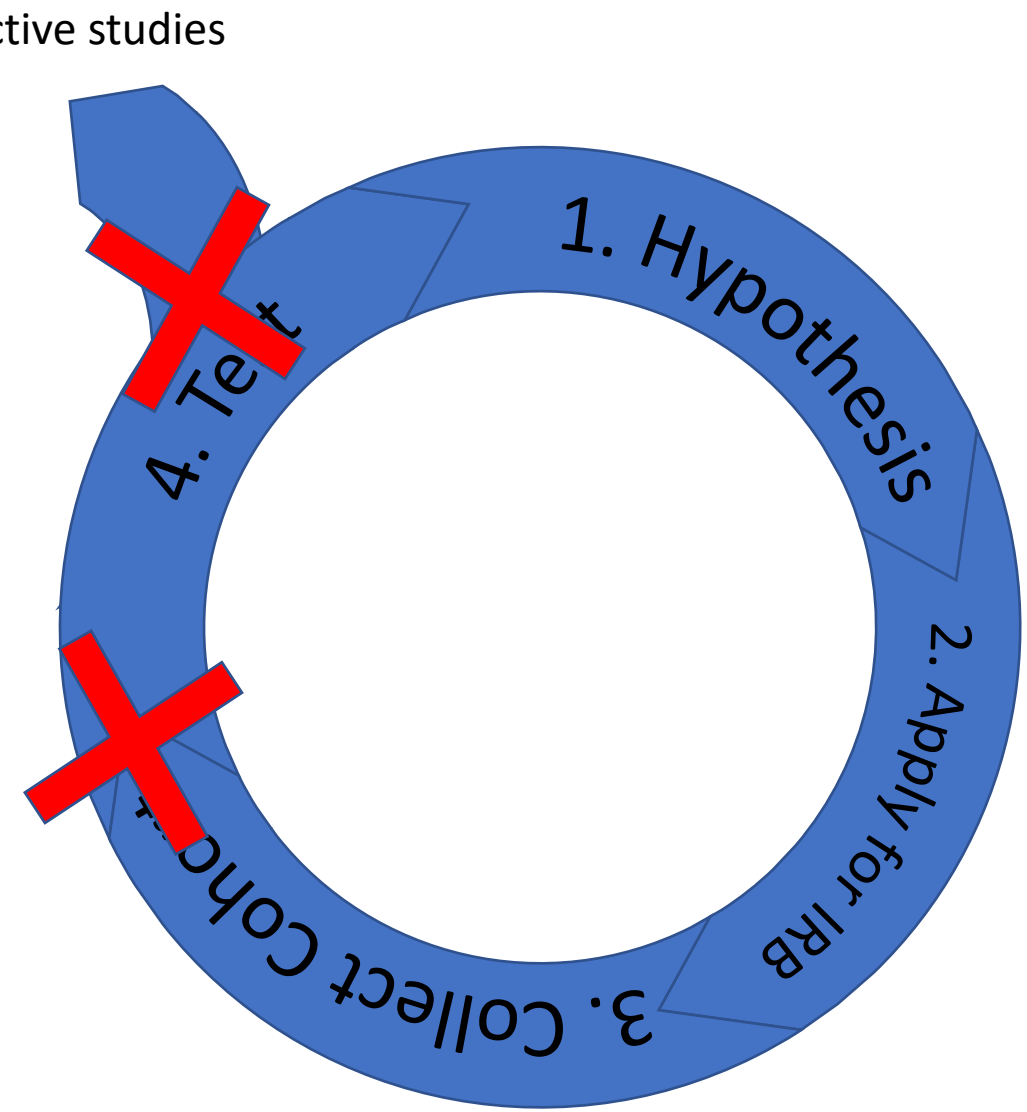
Real-world data for oncology questions big and small

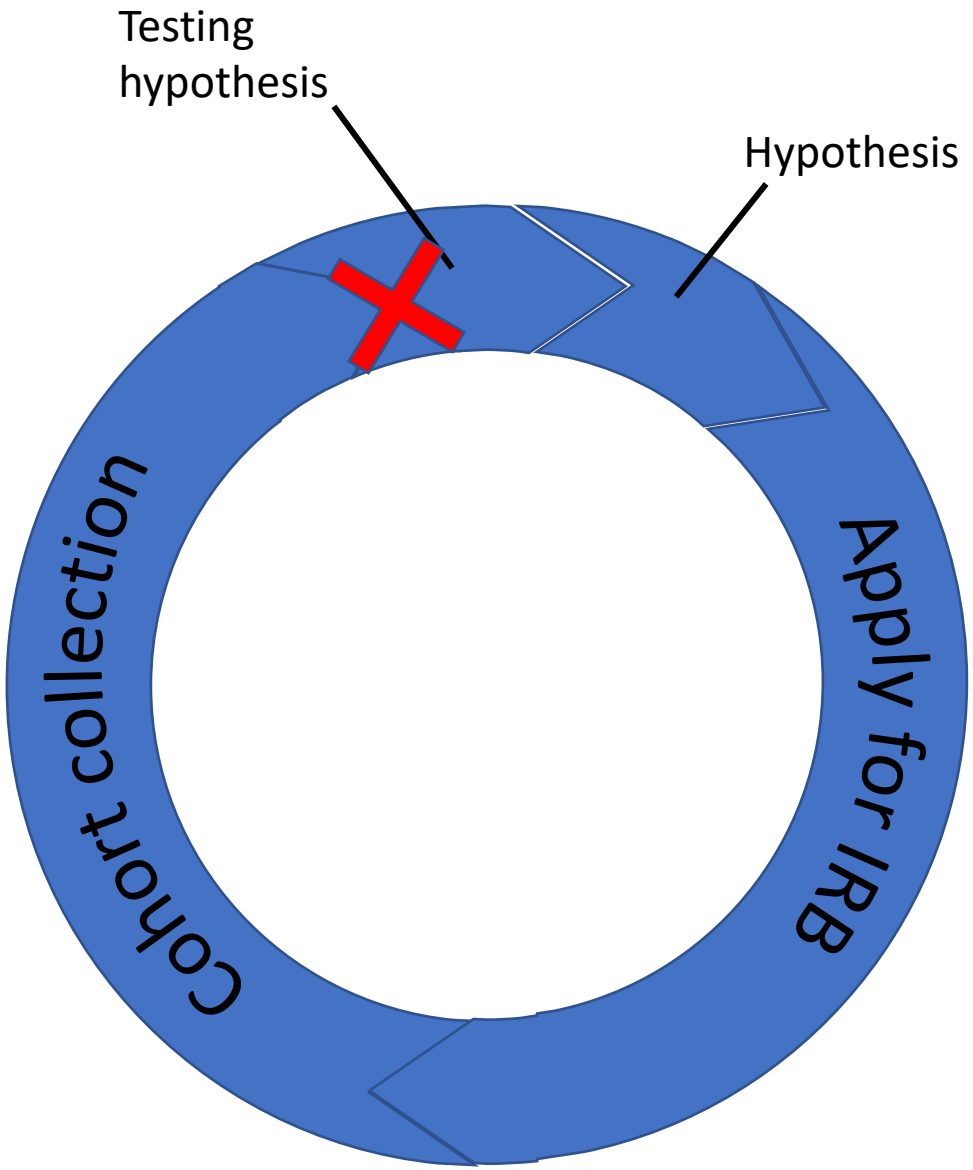
Travis Zack MD, PhD

Hypothesis testing in retrospective clinical data analysis



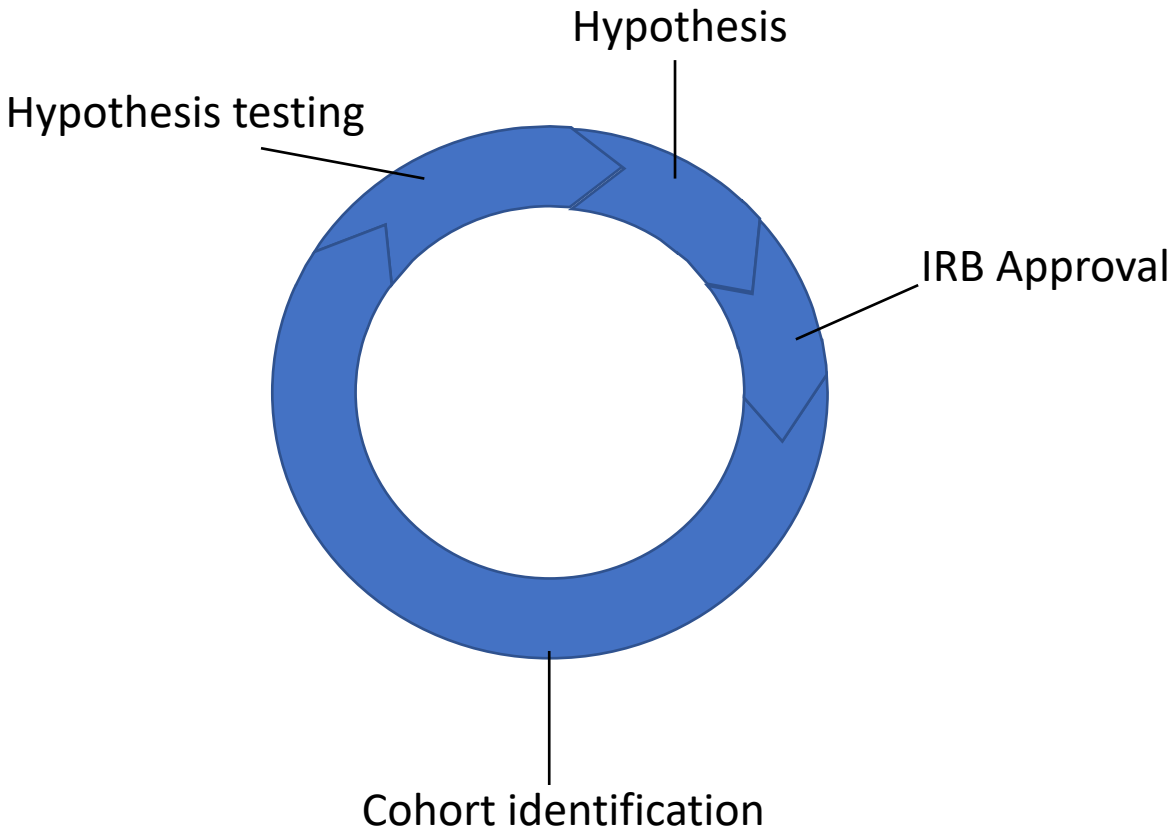
Research cycle for retrospective clinical data analysis

- In practice, iterative hypothesis testing in clinical data can be a arduous process
- Majority of time spent is on bureaucracy and data-wrangling
- Ideally, this would be minimized to free researchers to think critically



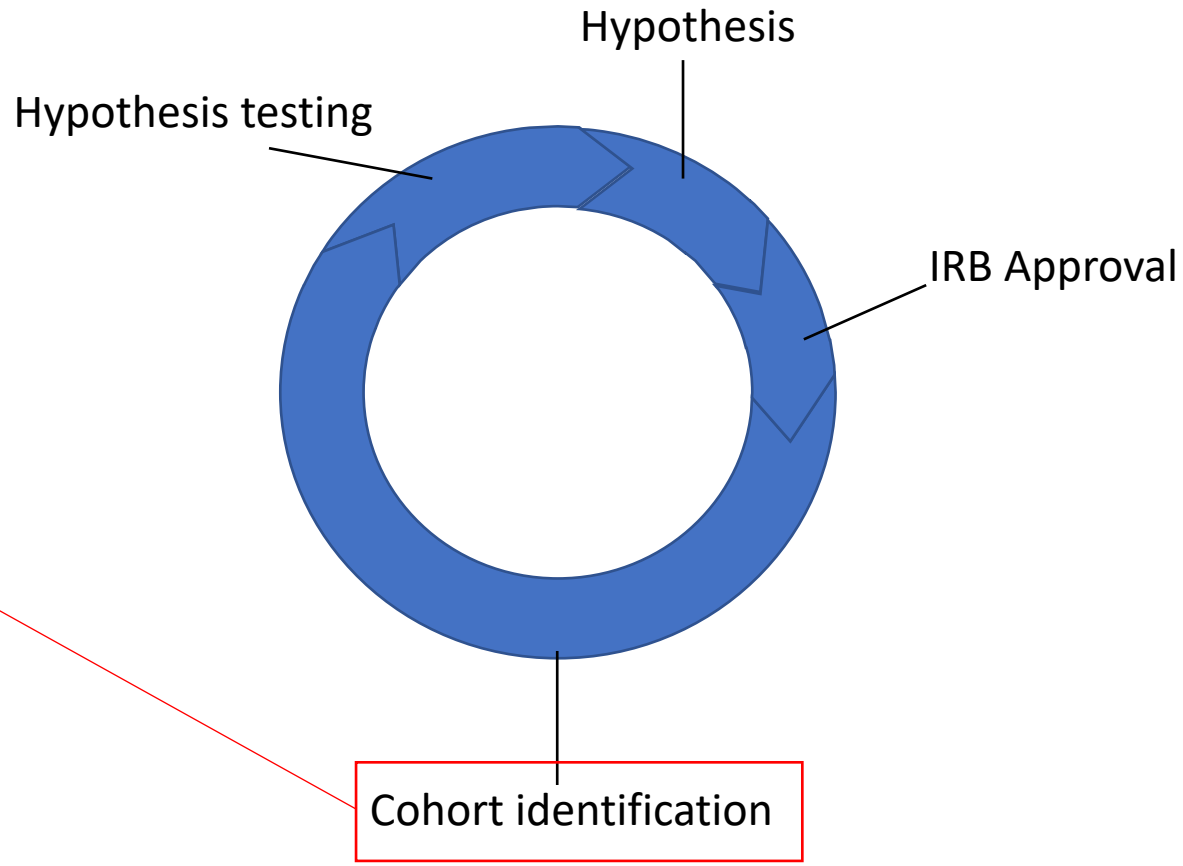
Research cycle for retrospective clinical data analysis

- Information commons decreases the time between iterations
- However, cohort identification for cancer hypotheses remains challenging



Research cycle for retrospective clinical data analysis

- Information commons decreases the time between iterations
- However, cohort identification for cancer hypotheses remains challenging



Exploring incidence and biochemistry of rare malignancies using real world clinical data

Exploring incidence and biochemistry of rare malignancies using real world clinical data

Josh Gordon

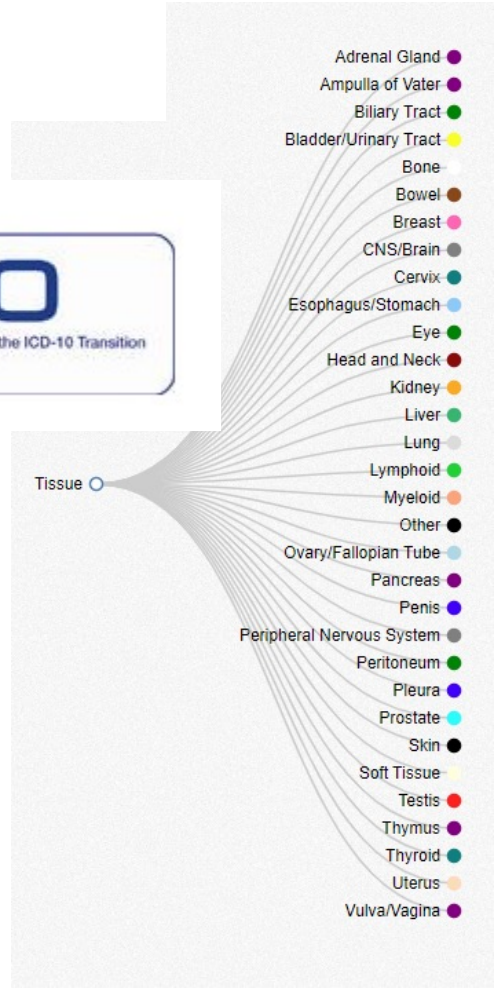
FCF

Fibrolamellar Cancer Foundation

Kurt Losert

Trade-offs in utilization of Standard Classification schema

- Electronic medical records and insurance billing attempt to discretize human disease in a number of different ways
- The most common way to do this is through “codes” or domain ontologies
- Allows for categorization at the expense of information loss.
- By design, not use-case specific

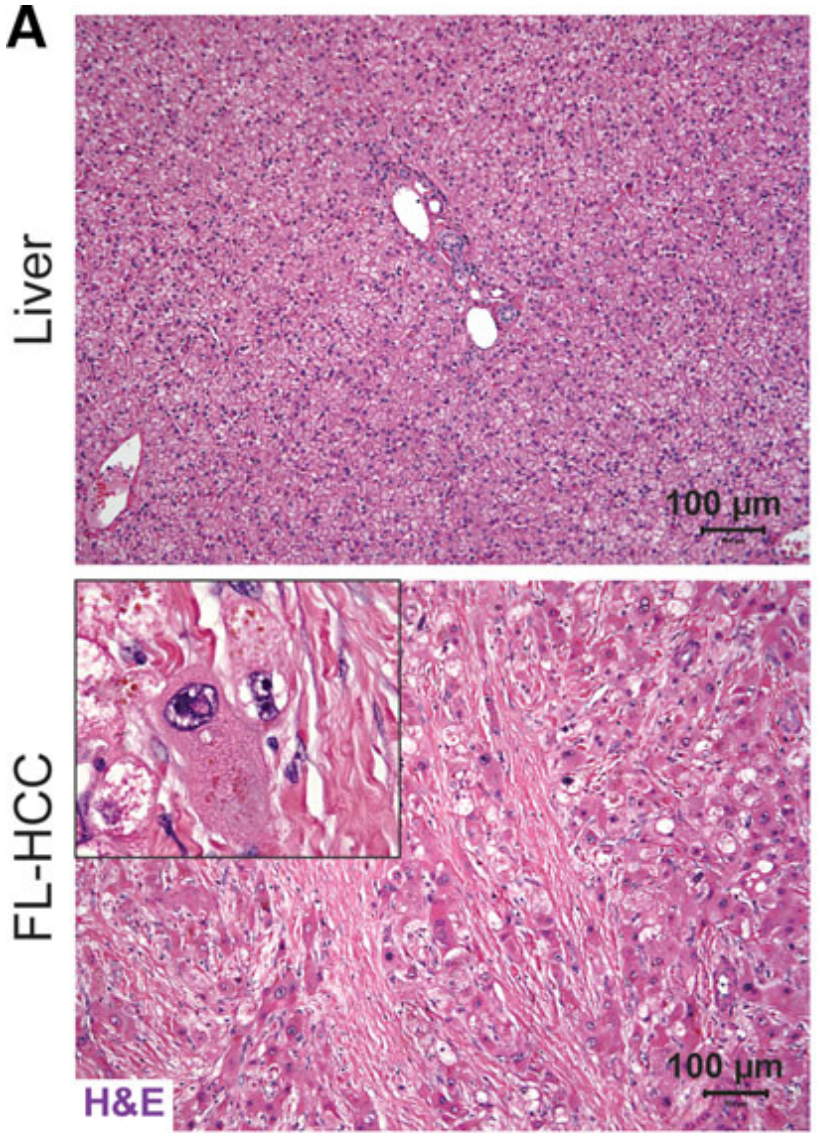


Fibrolamellar Hepatocellular Carcinoma (FLC)

SEER:

- Incidence 0.02/100,000
- Average age 22 years, but with a second group at 70-74 years

The SEER incidence rates and age distributions do not match expectations of clinical experts



Xu et al., *Hum Mol. Genet.* 2013

Ang et al., *Gastrointestinal Cancer Res.* 2013

FLC does occurs in otherwise healthy livers, and by unique molecular mechanism

Most recent SEER estimates:

- Incidence 0.02/100,000 (60-80 cases/year)
- Average age 22 years, but with a second group at 70-74 years

The SEER incidence rates and age distributions do not match expectations of clinical experts

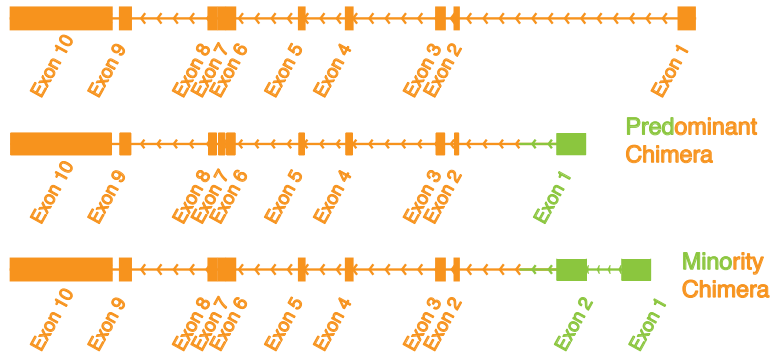
El-Serag, H. B. & Davila, J. A. *Hepatology* **39**, 798–803 (2004).

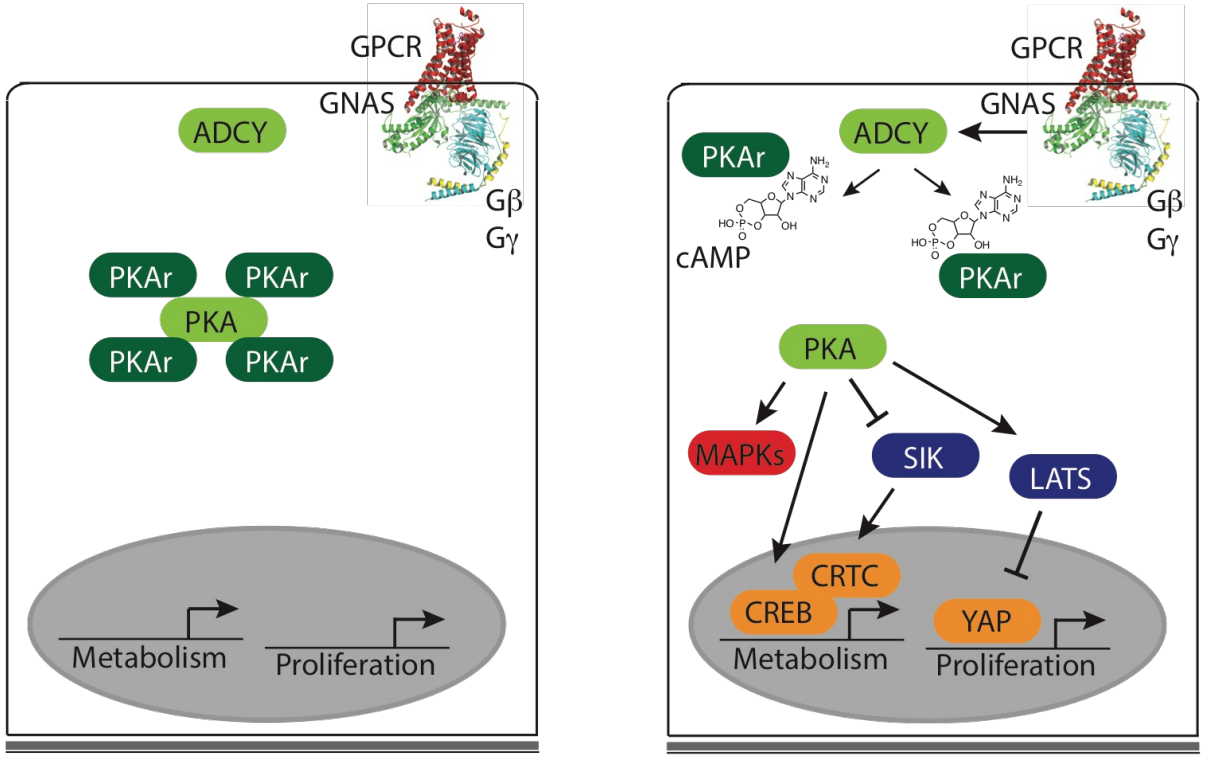
FLC does occurs in otherwise healthy livers, and by unique molecular mechanism

- Most liver cancer occurs in chronically inflamed liver
- FLC typically occurs in young patients without comorbities.
- Unique molecular event that has biochemical ramifications but poorly understood

Detection of a Recurrent *DNAJB1-PRKACA* Chimeric Transcript in Fibrolamellar Hepatocellular Carcinoma

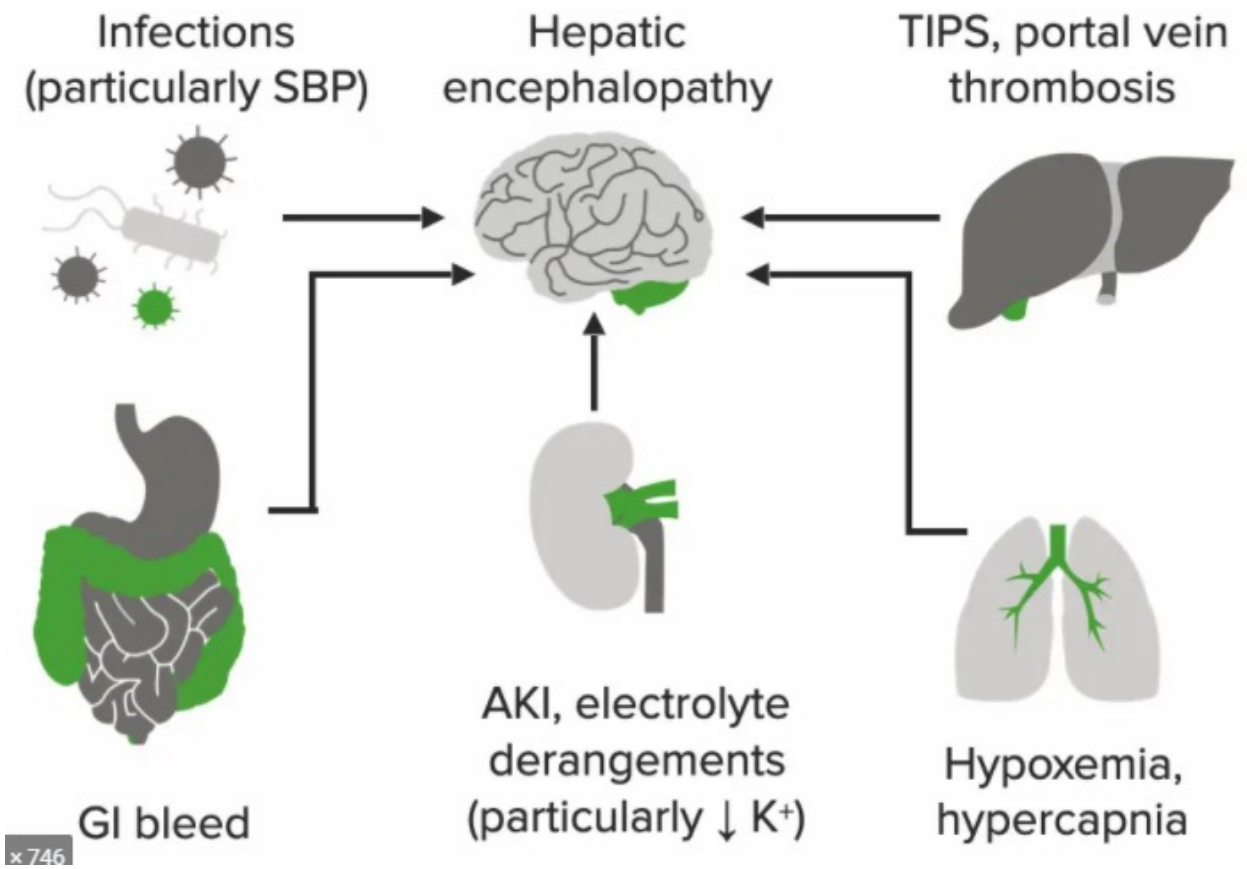
Joshua N. Honeyman,[1,2]* Elana P. Simon,[1,3]* Nicolas Robine,[4]* Rachel Chiaroni-Clarke,[1]





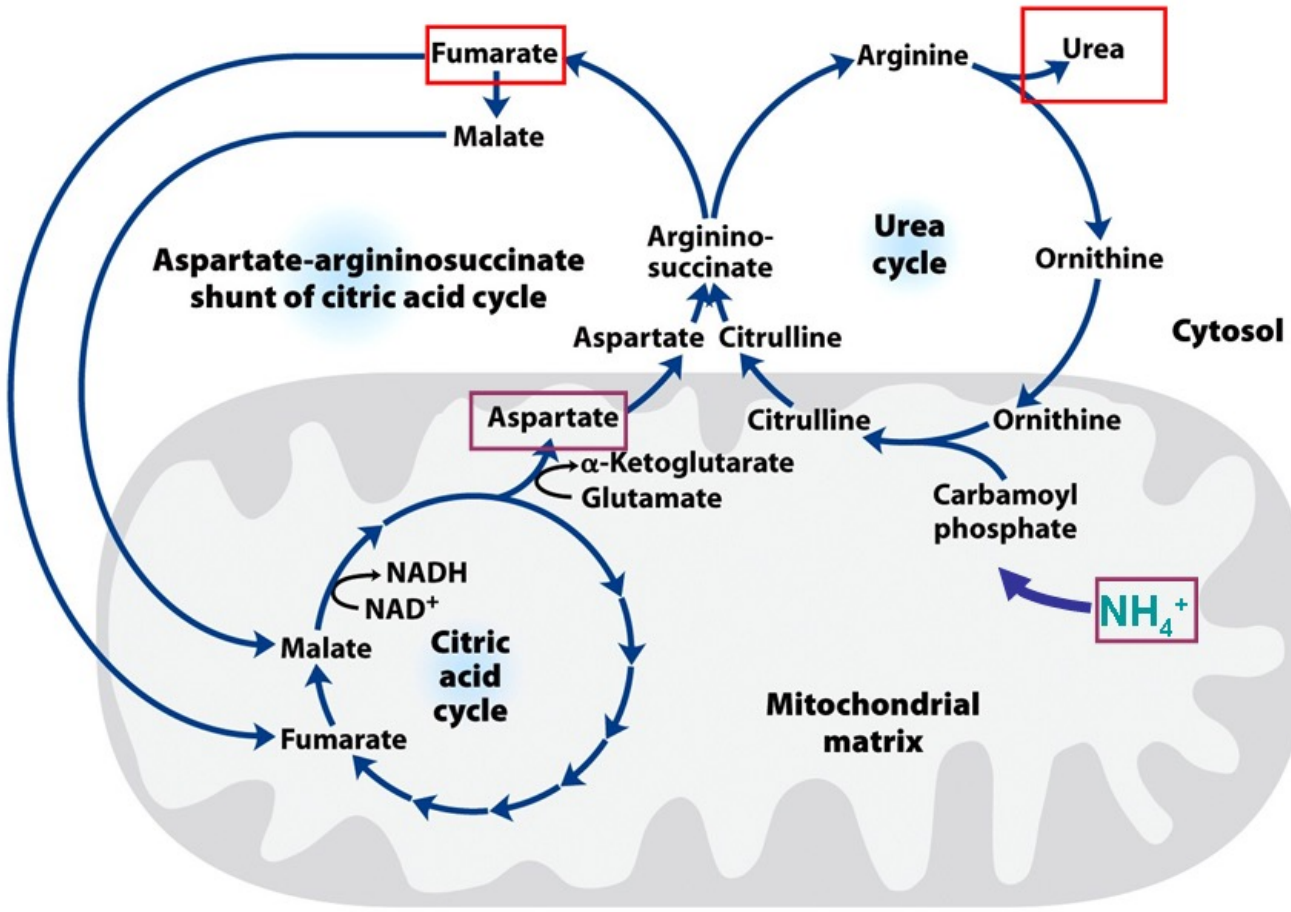
Hyperammonemia and FLC

- Hyperammonemia is a major complication of Liver insufficiency
- About 30-70% of patients with cirrhosis are thought to experience HE at some point
  - Readmission rates for patients with previous HE extremely high
- Complications relating to HE direct costs are estimated at ~5k-50k/patient/year
- Despite having healthy livers, FLC patients have been noted to suffer from severe cases of hyperammonemia



Pharmacoeconomics. 2018; 36(7): 809–822.

Hyperammonemia in FLC: Biochemical hypothesis

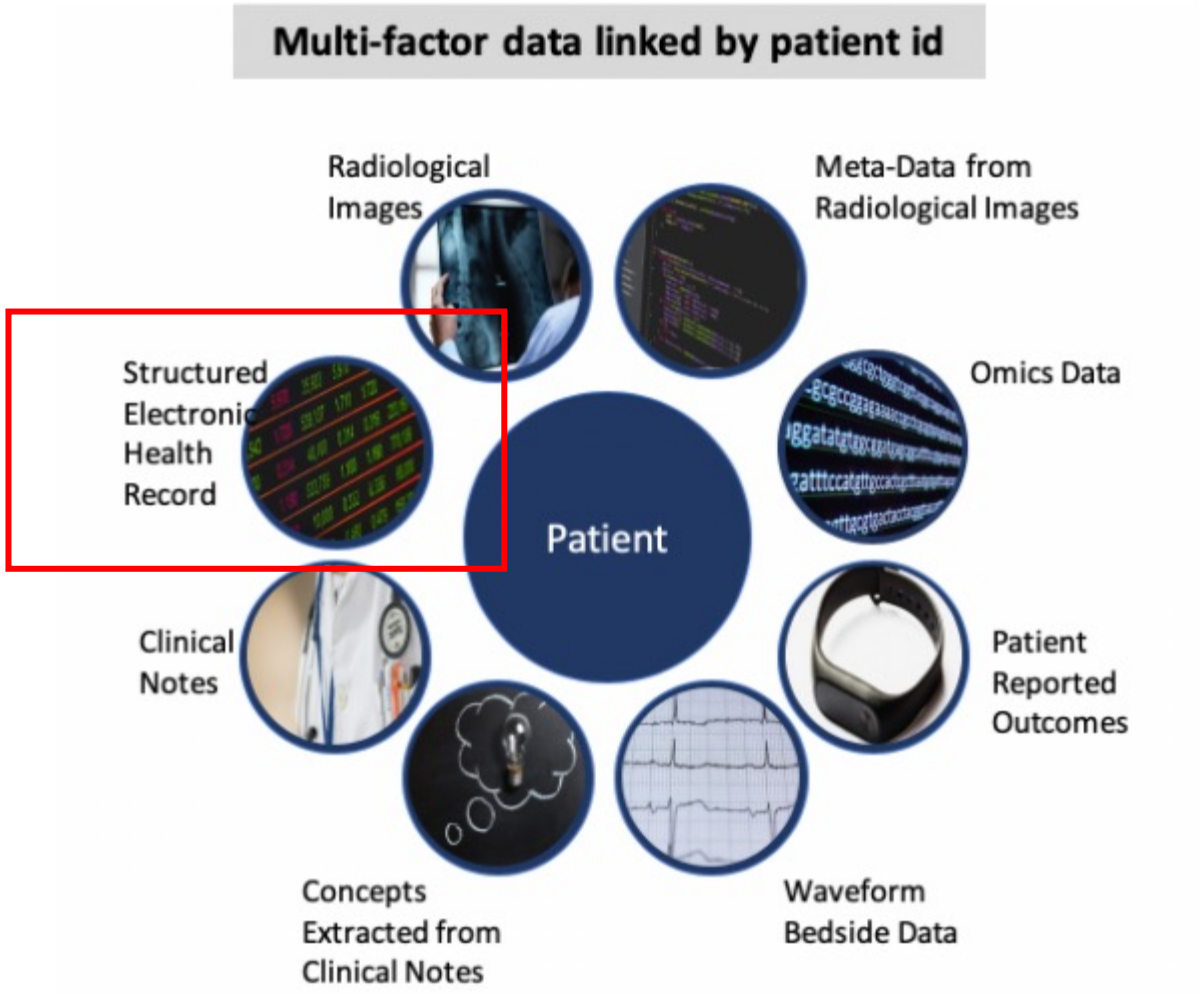


- Waste management: processes excess nitrogen from amino acids
- Required for gluconeogenesis – processes AA into carbon for the citric acid cycle
- **PKA** inhibits glycolysis, requiring AA breakdown for αKG to enter the TCA cycle

Standard Classification schema can misrepresent incidences of rare diseases

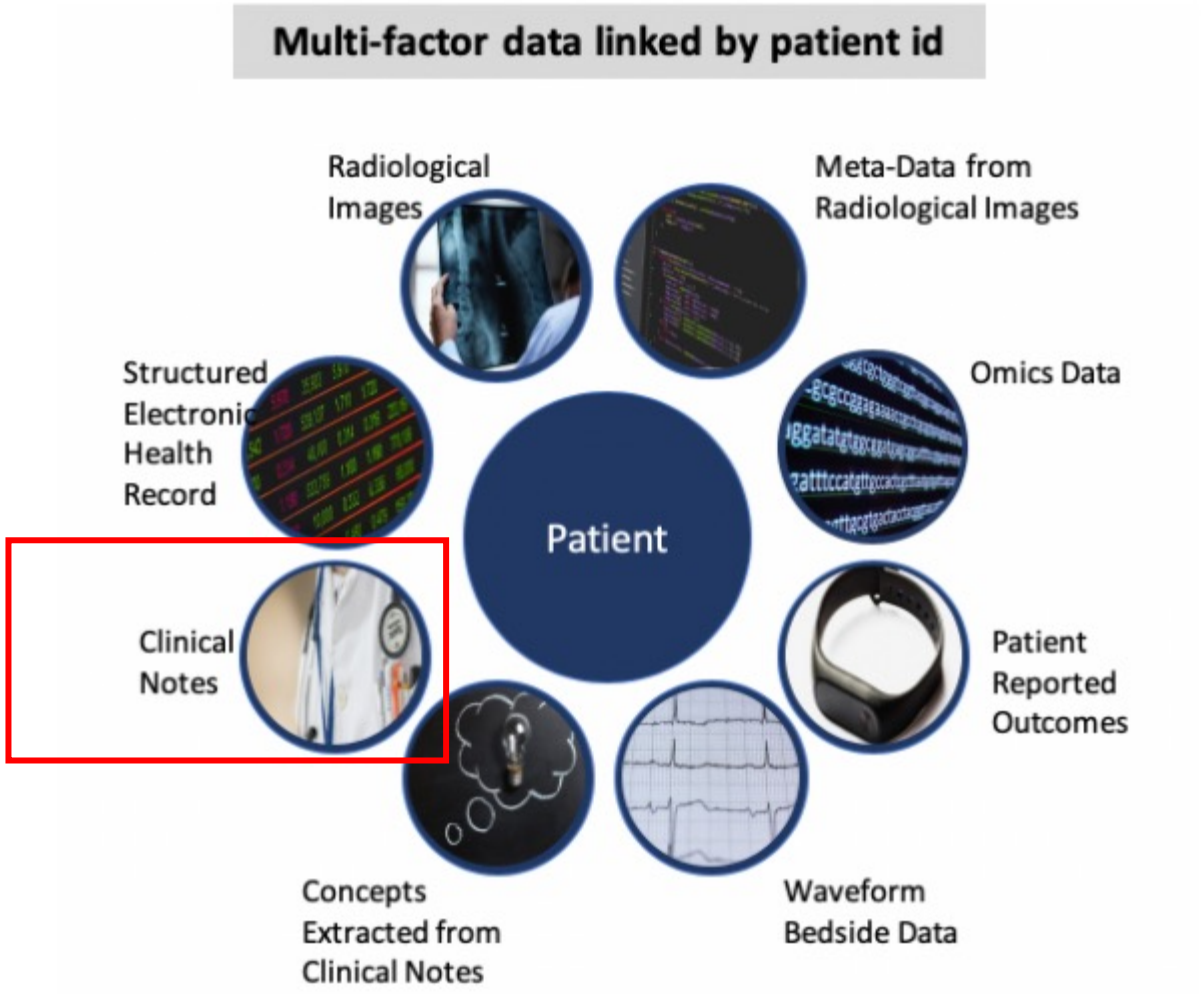
- Most common classification schema used nationally is ICD-9/10.
- Categorizes many cancers by organ site, rather than histopathology or molecular characteristics
- Does not contain individual categories for many rare diagnoses.

Identifying FLC patients at UCSF



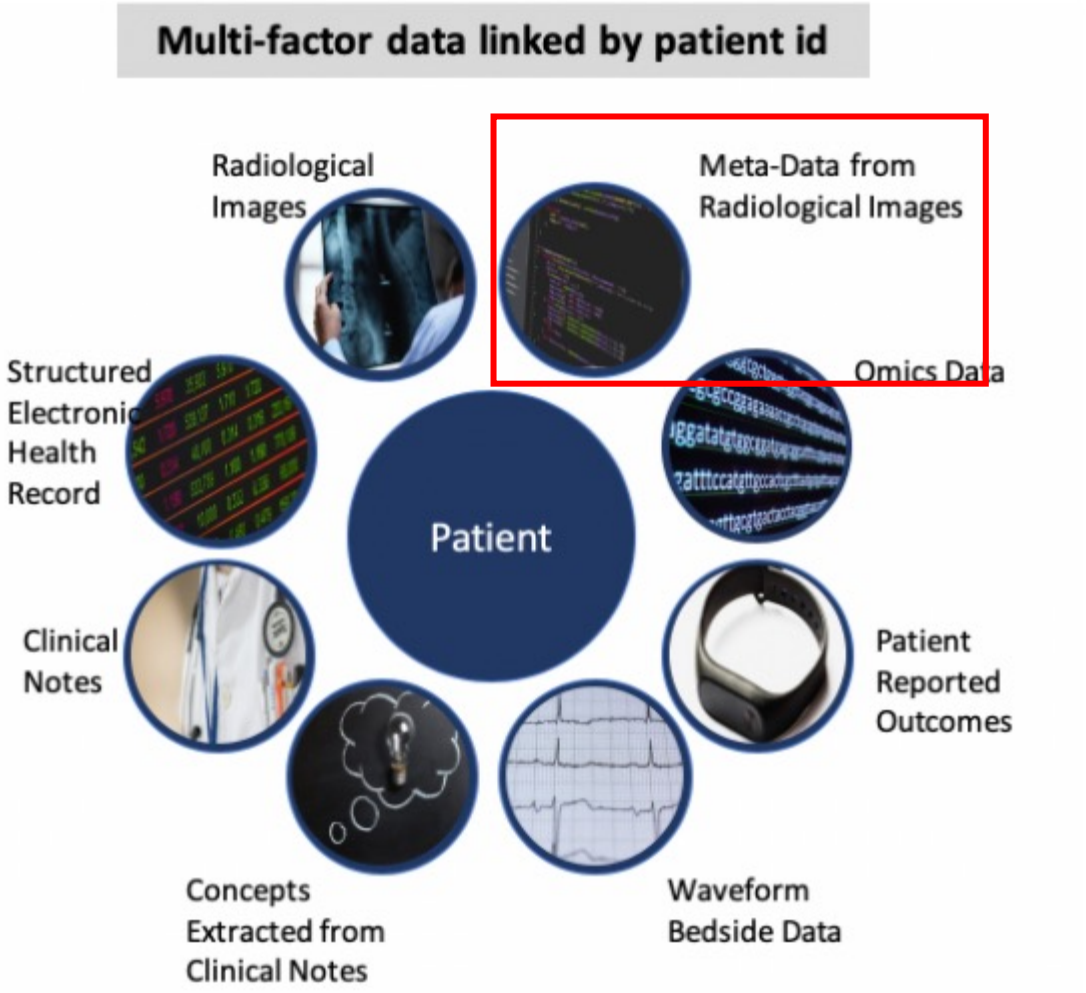
- EMR-billing codes (SNOMED)

Identifying FLC patients at UCSF



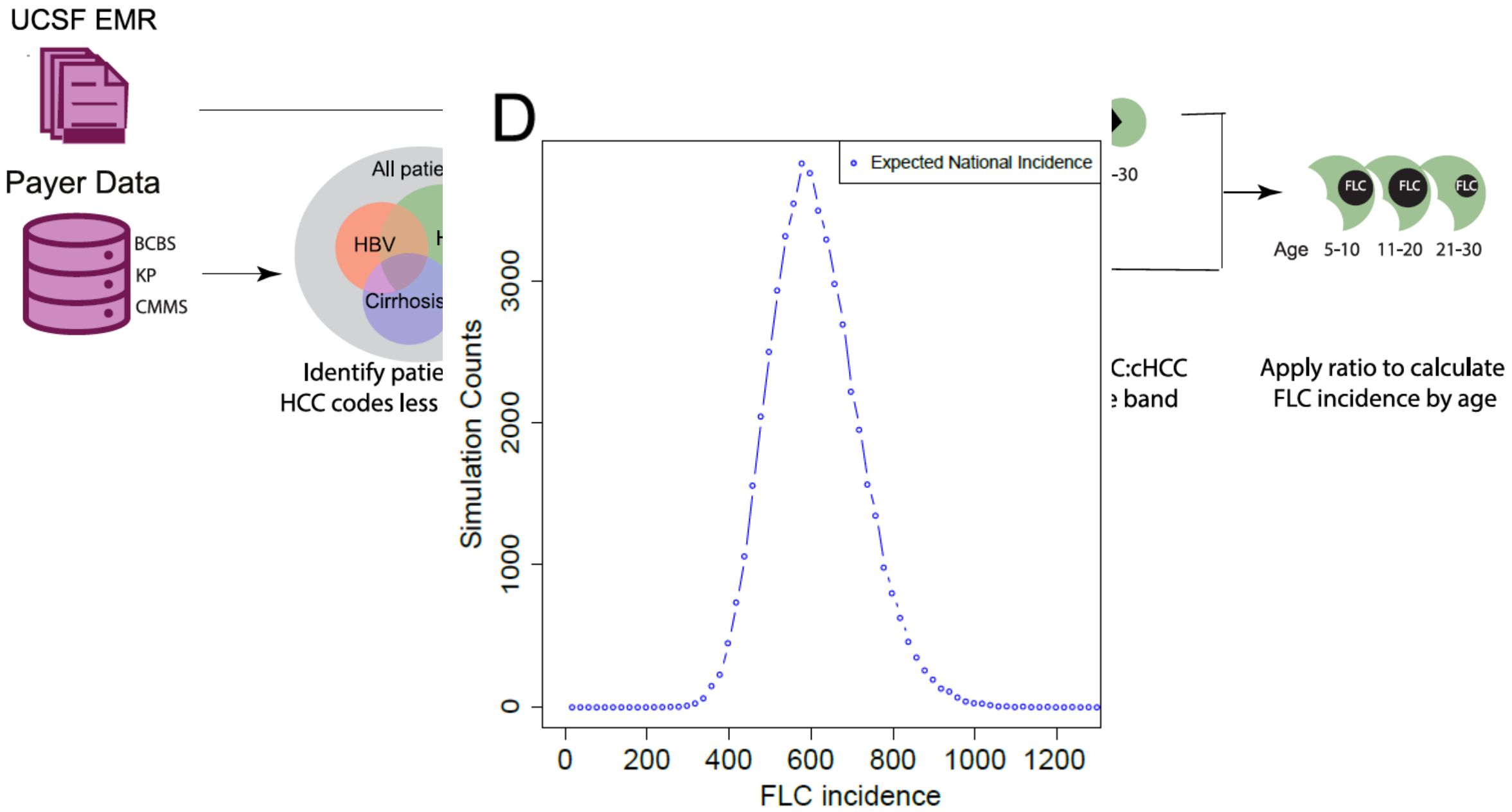
- EMR-billing codes (SNOMED)
- Term extraction and negation filtering of
  - Oncologist notes
  - Surgical notes
  - Pathology reports

Identifying FLC patients at UCSF



- EMR-billing codes (SNOMED)
- Term extraction and negation filtering of
  - Oncologist notes
  - Surgical notes
  - Pathology reports
- Radiology report identification

A


UCSF EMR
Payer Data
BCBS
KP
CMMS
All patie
HBV
Cirrhosis
Identify patie
HCC codes less
-30
C:cHCC
e band
FLC
FLC
FLC
Age 5-10 11-20 21-30
Apply ratio to calculate
FLC incidence by age

D

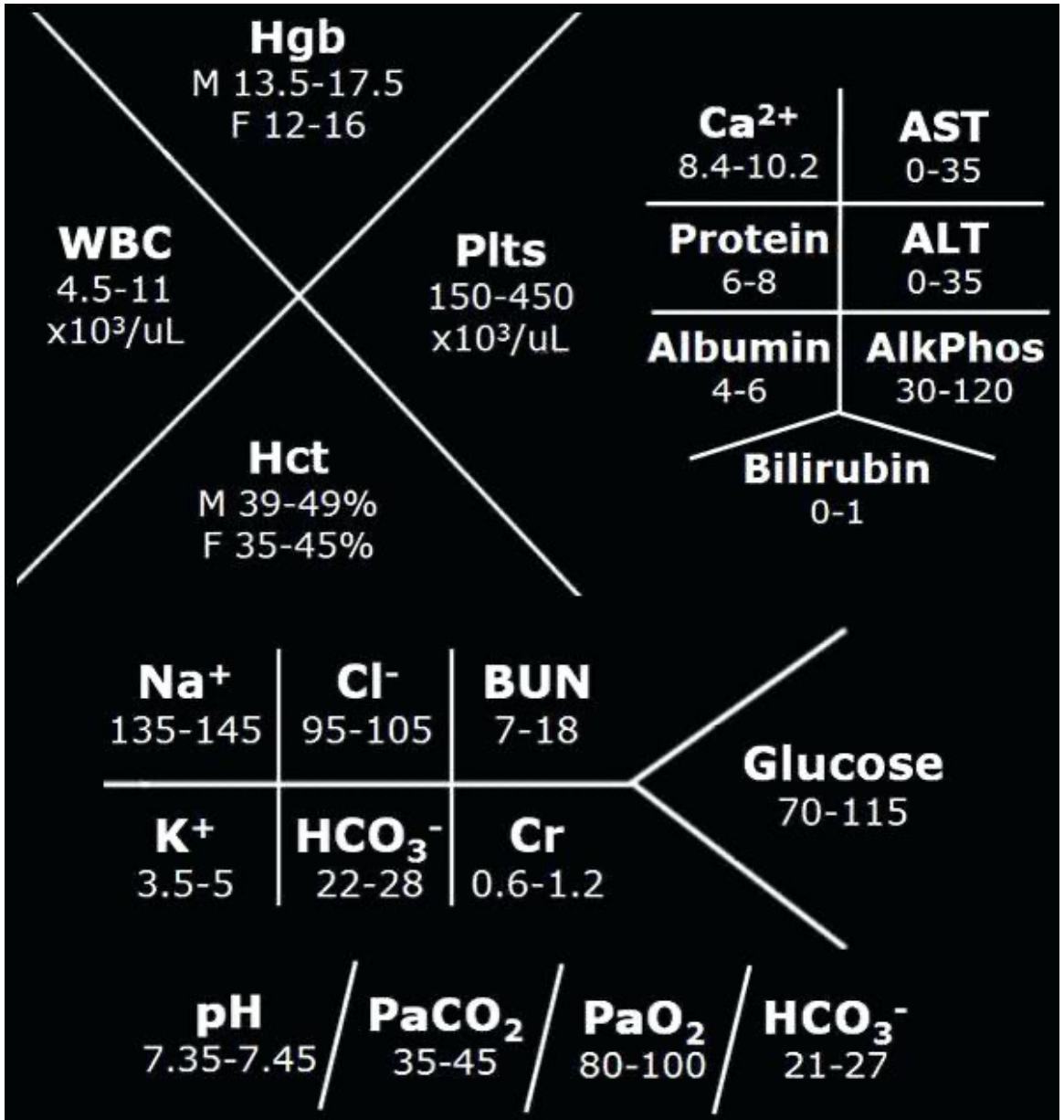
Expected National Incidence
Simulation Counts
0
1000
2000
3000
0
200
400
600
800
1000
1200
FLC incidence

Using real-world data to elucidate mechanism of hyperammonemia

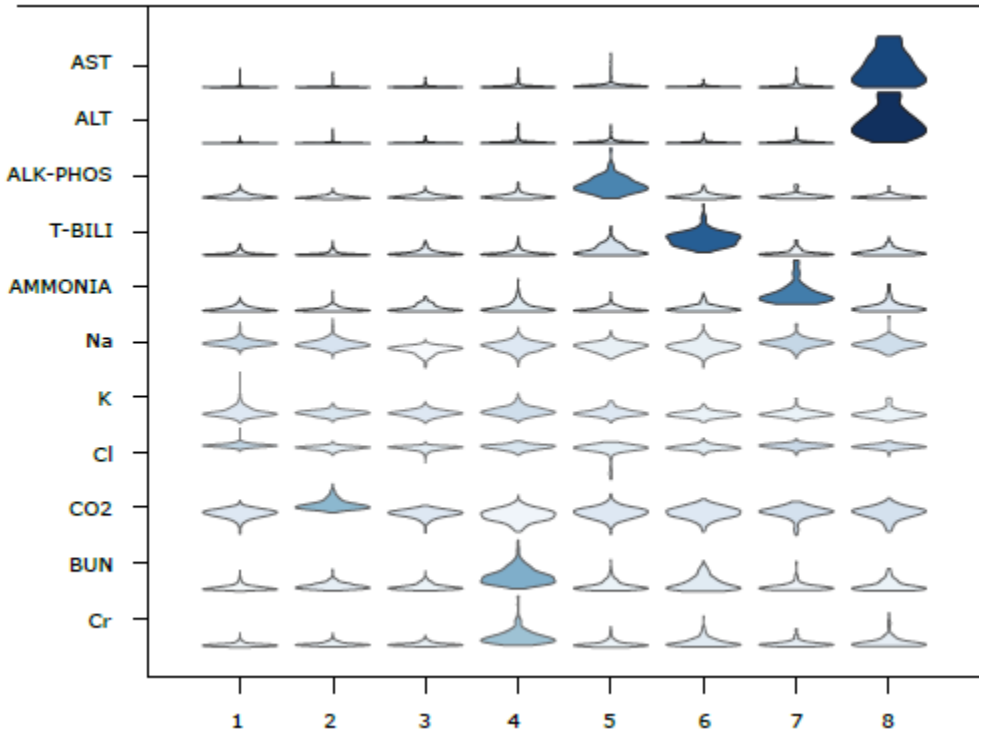
The gene mutated in FLC is responsible for Ammonia metabolism

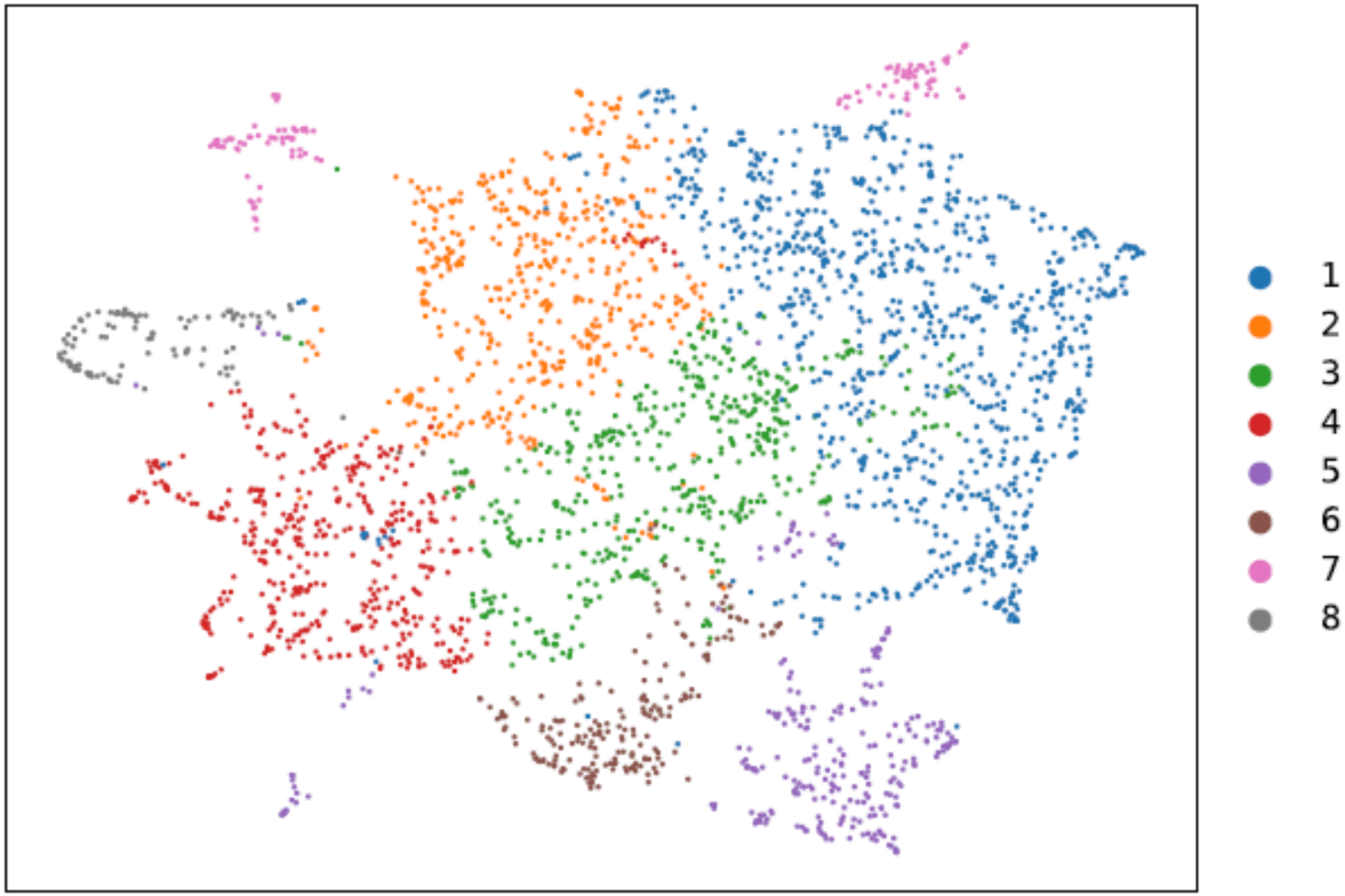
We hypothesized that, if the hyperammonemia in FLC is related to this mutation, the metabolic state of hyperammonemia patients with FLC may be more similar to patients with inborn defects of Ammonia metabolism, instead of broader liver failure.

To test this, we analyzed clinical laboratory data from each patient similar to gene expression analysis in scRNA

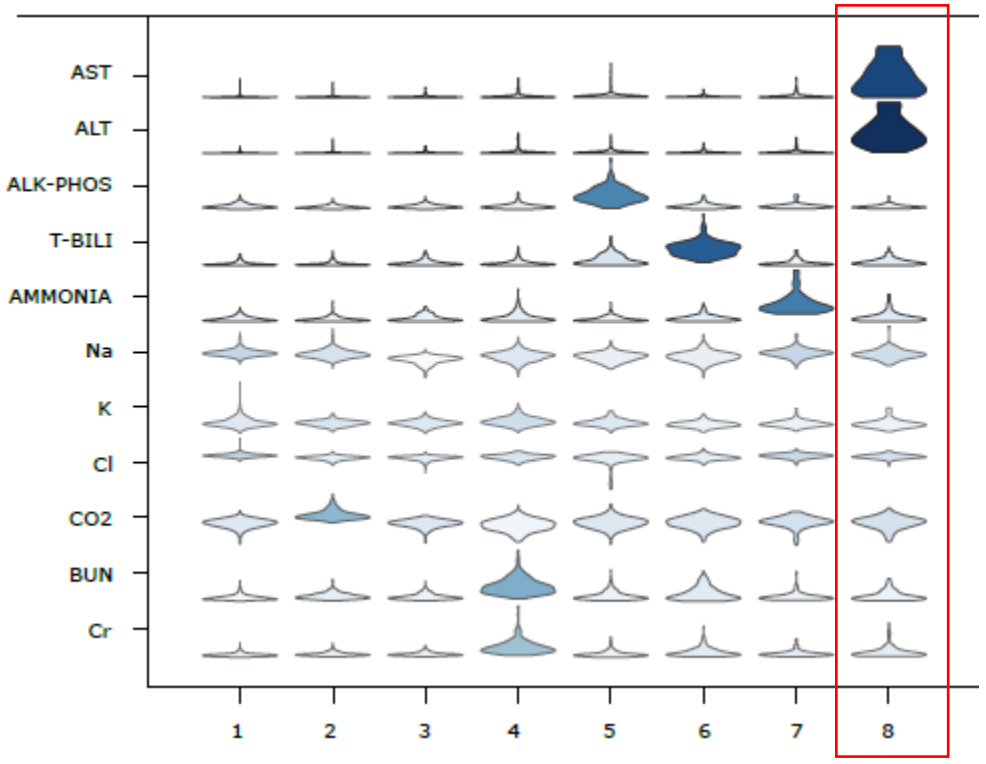


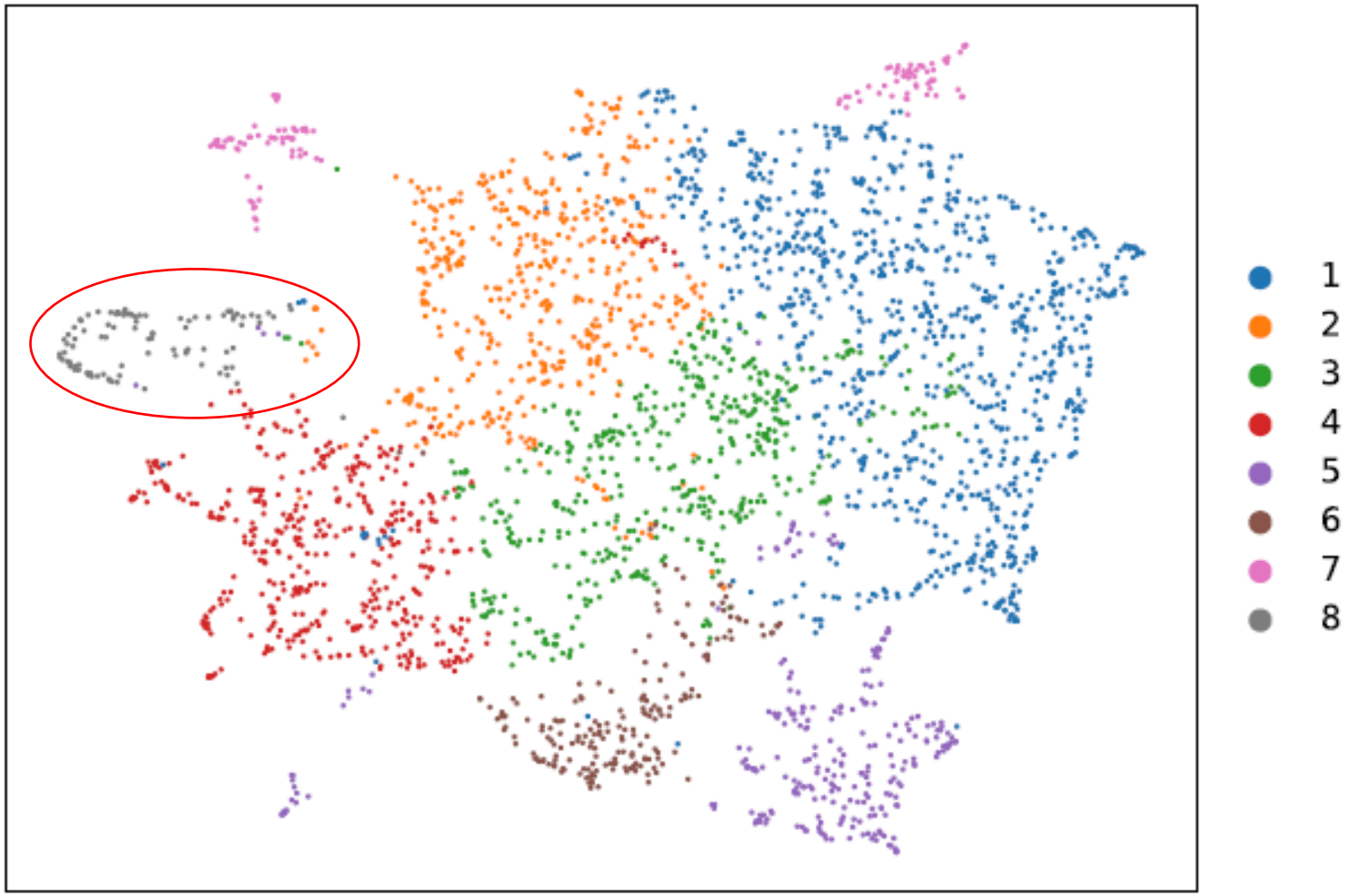
Metabolic lab data from hyperammonemia patients reveals eight distinct states





Metabolic lab data from hyperammonemia patients reveals eight distinct states



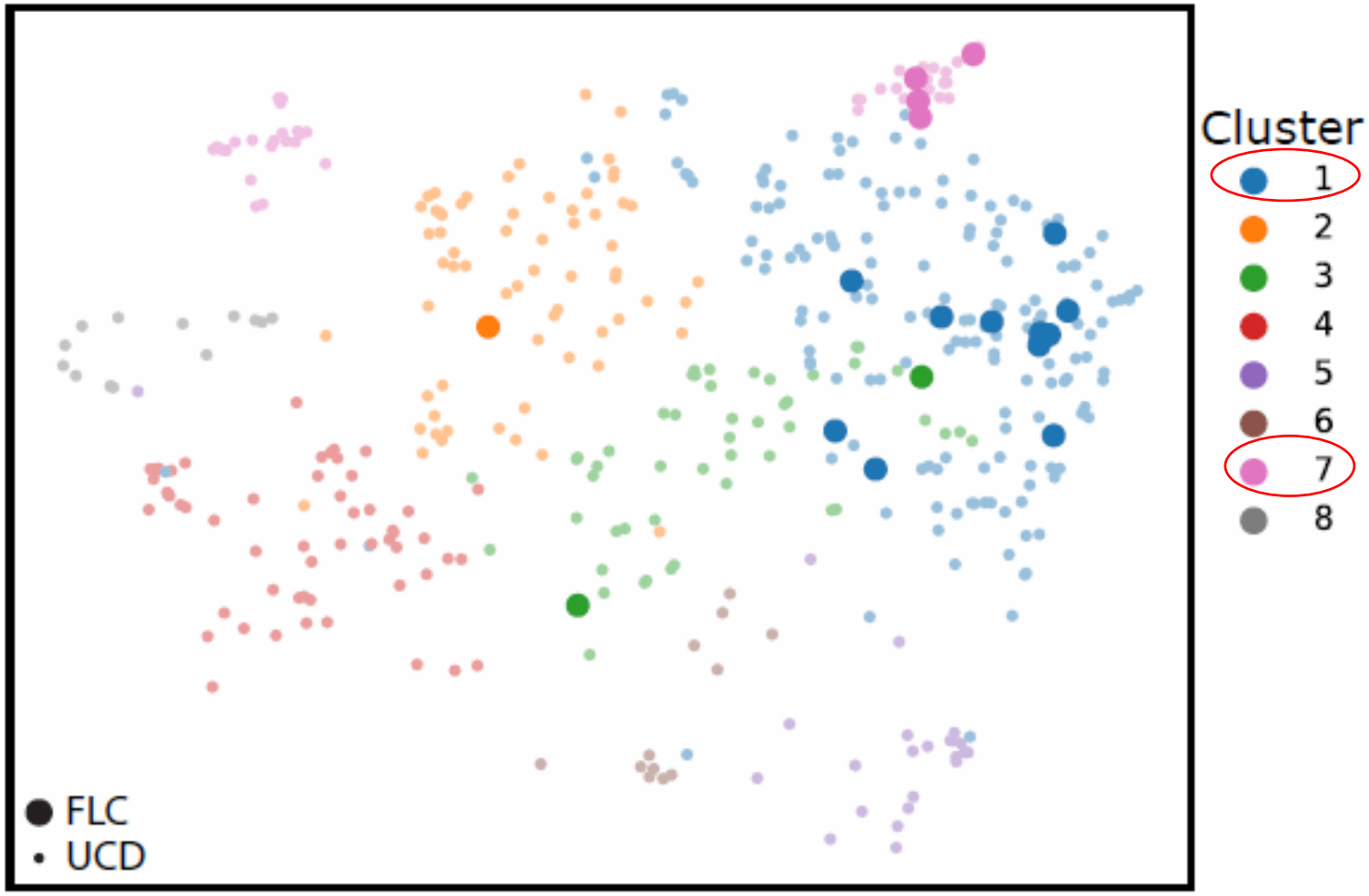


Metabolic state of FLC patient significantly clusters with patients with inborn errors of Urea Cycle

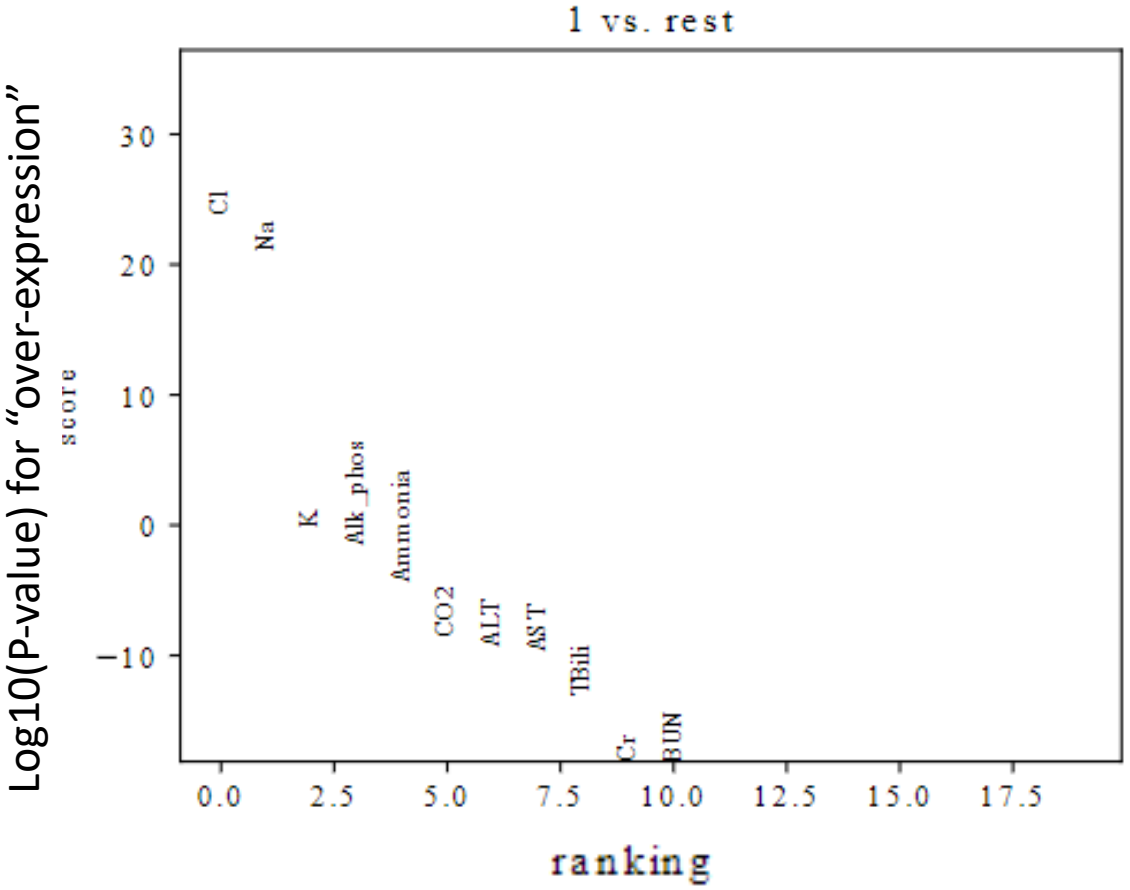
FLC patients were significantly enriched in two clusters: #1 and #7 ($p=0.0005$)

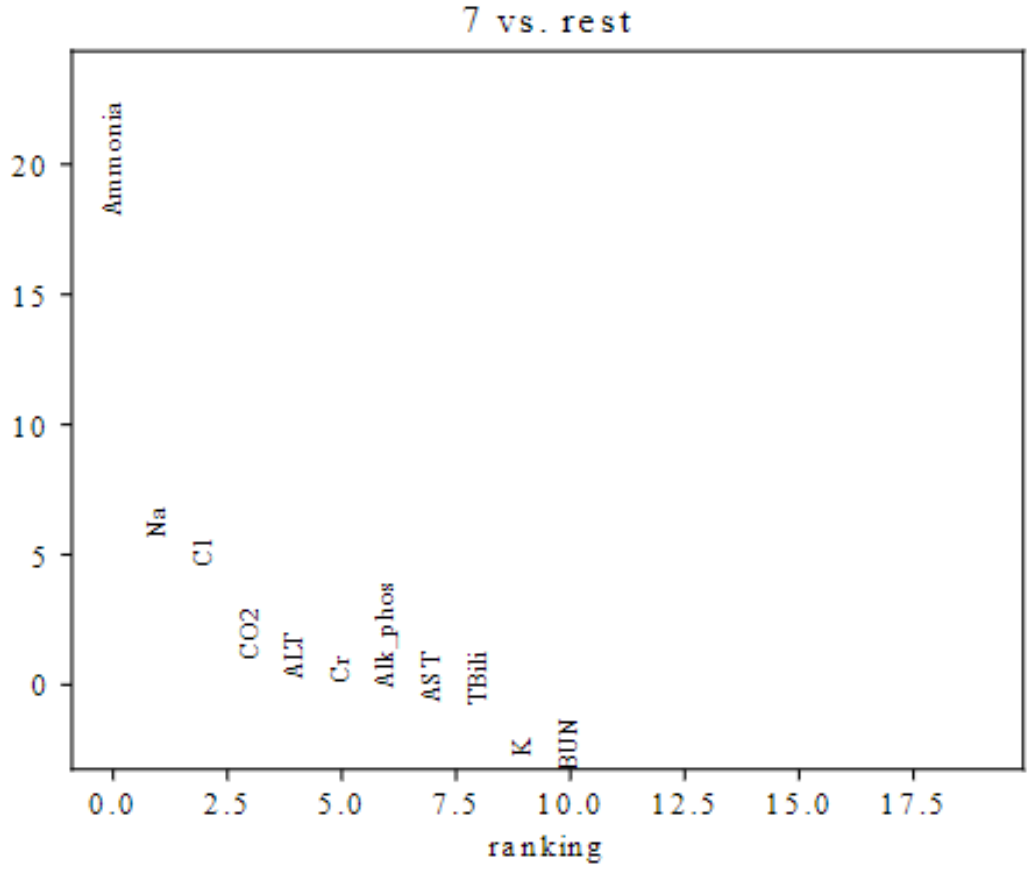
These same two groups were significantly enriched for patients with UCD ($p=3x10^{-10}$)

No such significance was seen for patients with classical hepatocellular carcinoma



Metabolic states of enriched for FLC patients: “Metabolically healthy” and “hyper”-hyperammonemia

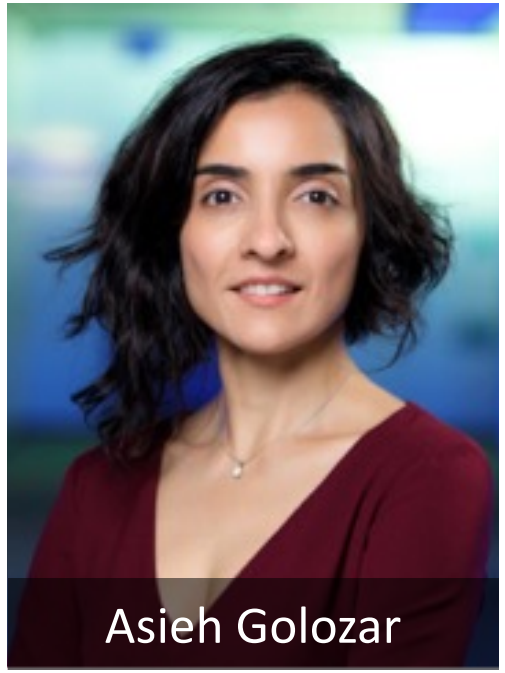




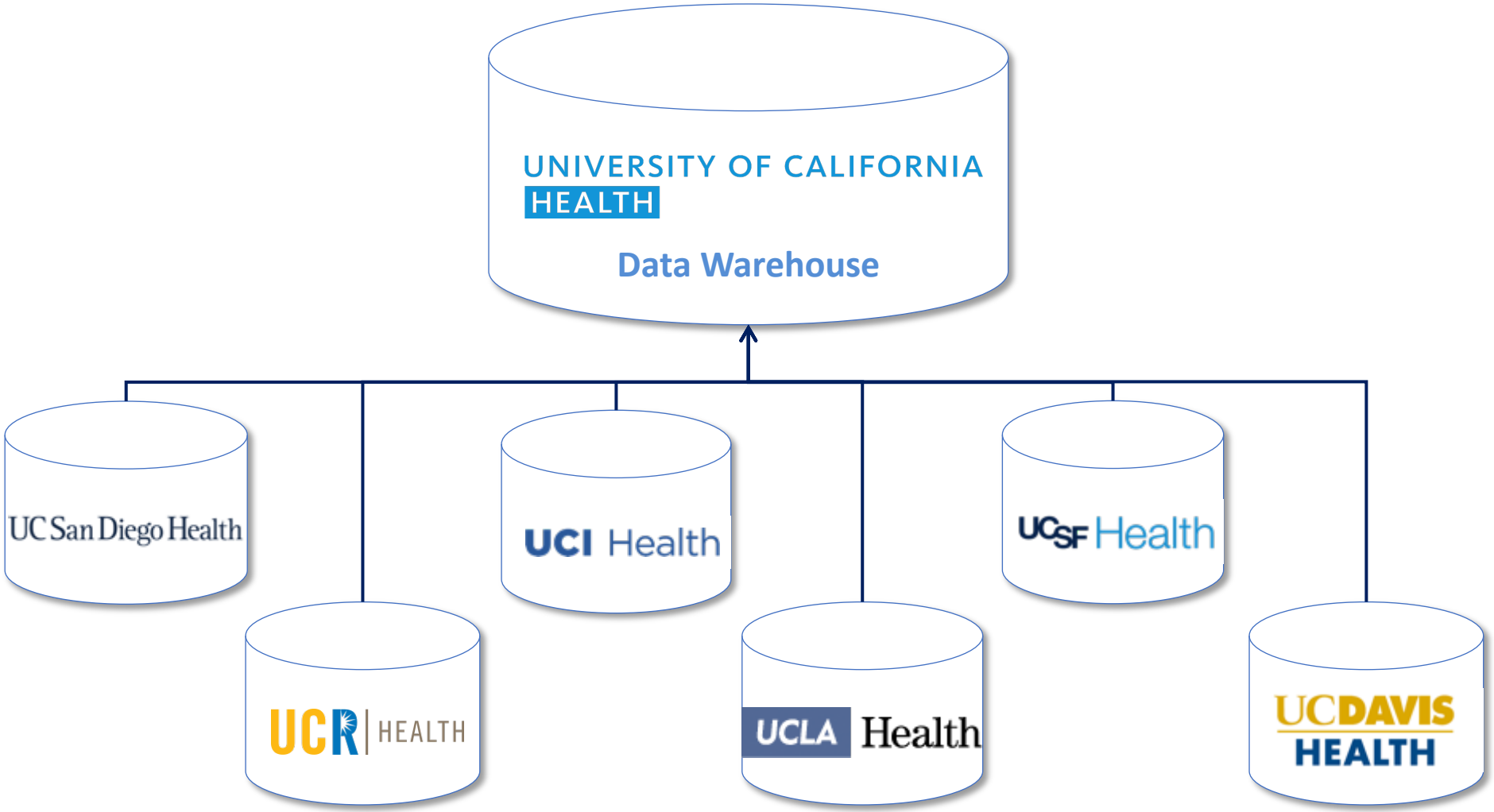
Accurate identification of oncologic treatment in real-world clinical data

OHDSI

HemOnc.org



Combining healthcare data from across the six University of California medical schools and systems



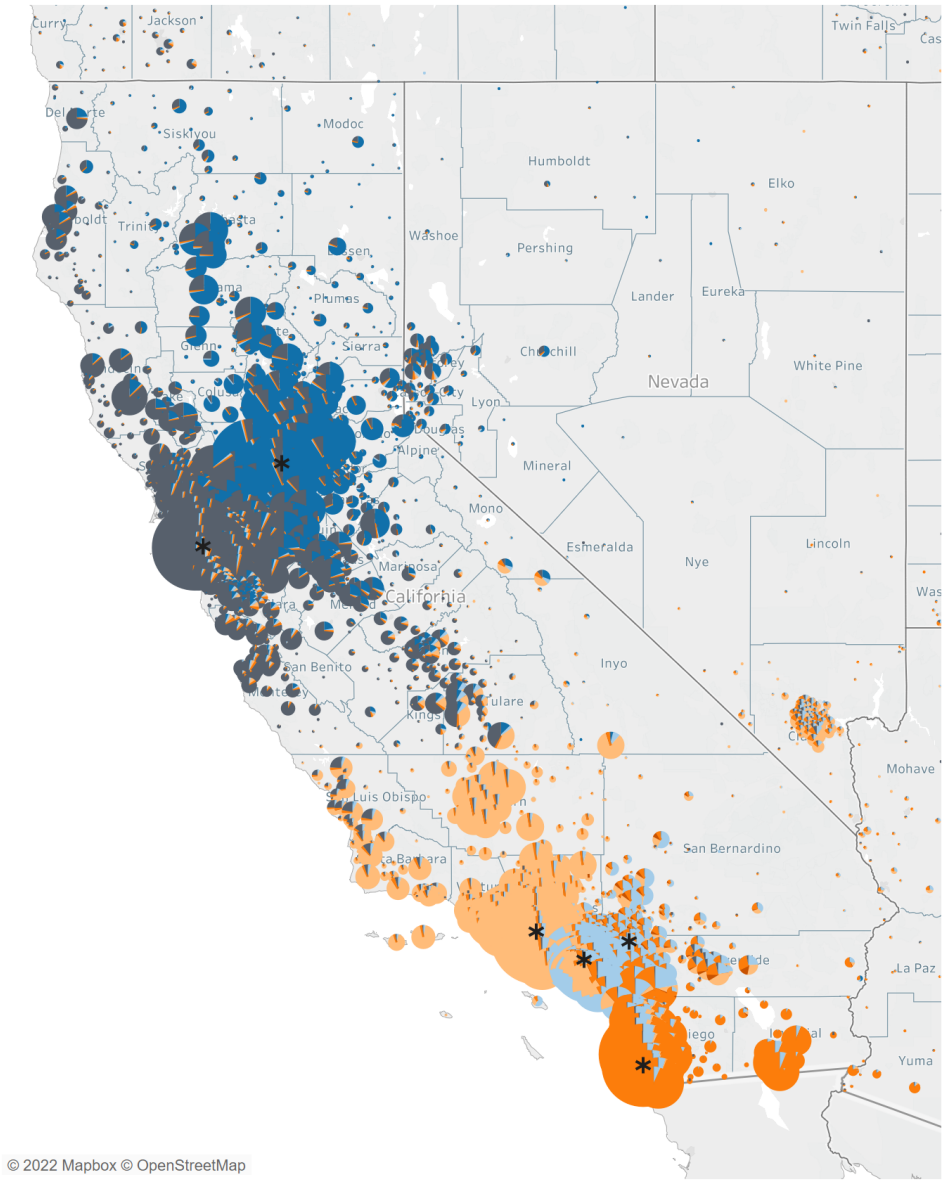
UC Health Data Analytics Platform

UNIVERSITY OF CALIFORNIA HEALTH

UC Health Patients (since January 2012)

Gender	ADI	Race	Current Age	Active?	PCP?	Ethnicity
All	All	All	All	All	All	All

Patient Volume by Home Zip Code



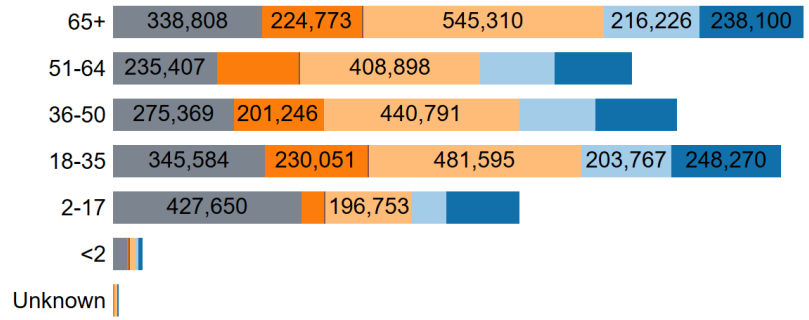
Most Recent Visit Date Range

January 2012 to December 2021
and Null values

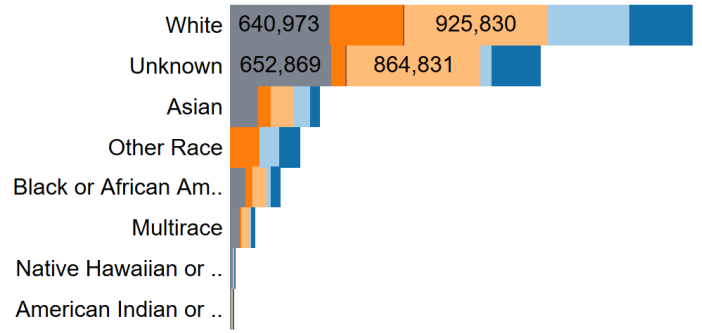
6,532,320 Total UC Health Patients

UCSF	UCSD	UCLA	UCI	UCD
1,654,208 (25.3%)	900,650 (13.8%)	2,092,970 (32.0%)	848,964 (13.0%)	1,020,154 (15.6%)

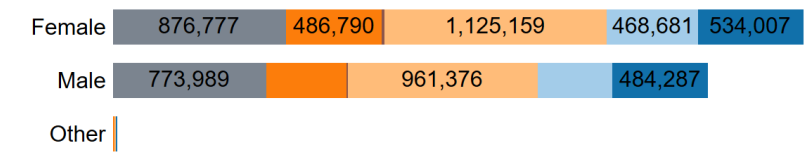
Age Group



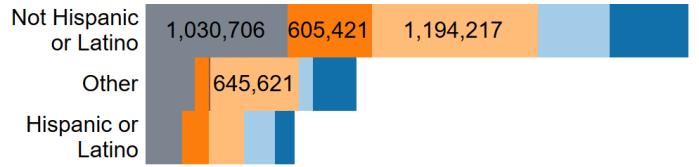
Race



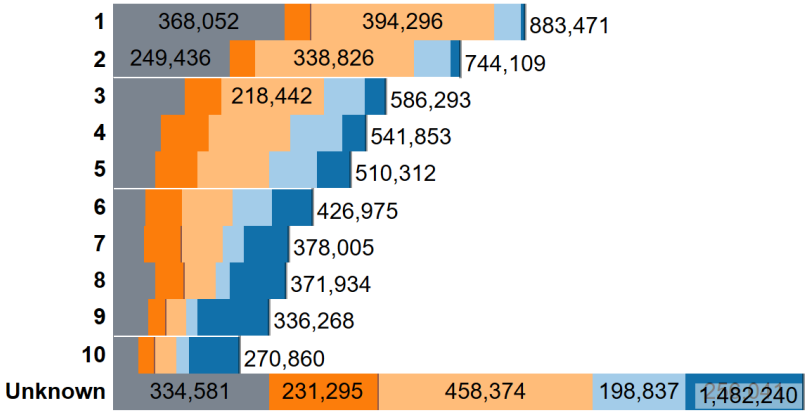
Gender



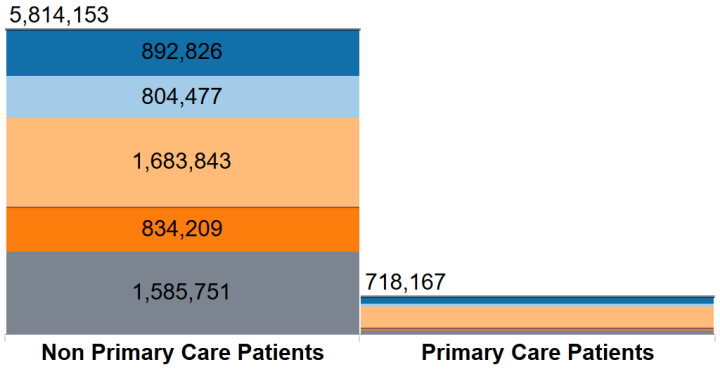
Ethnicity



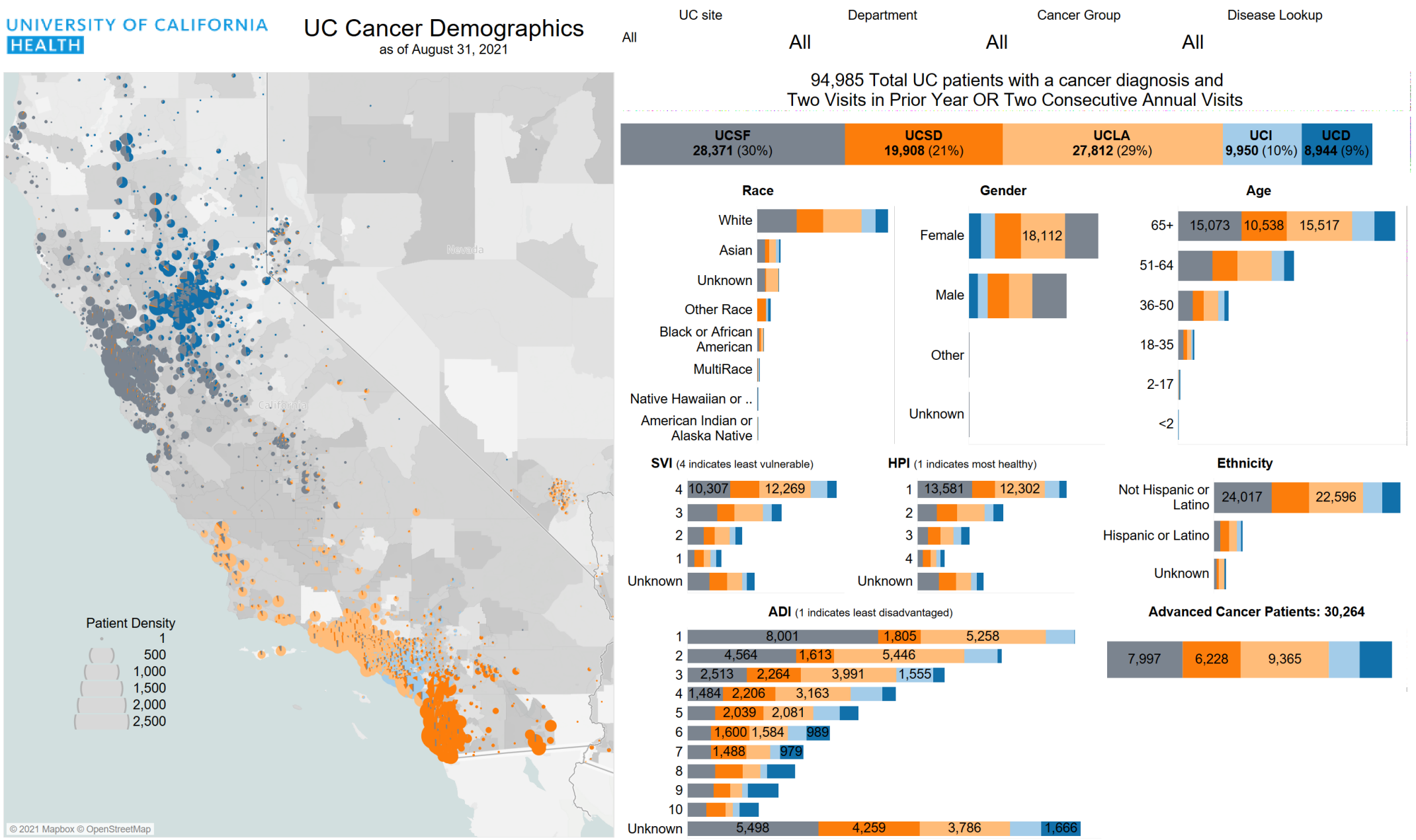
Area Deprivation Index



Primary Care Population

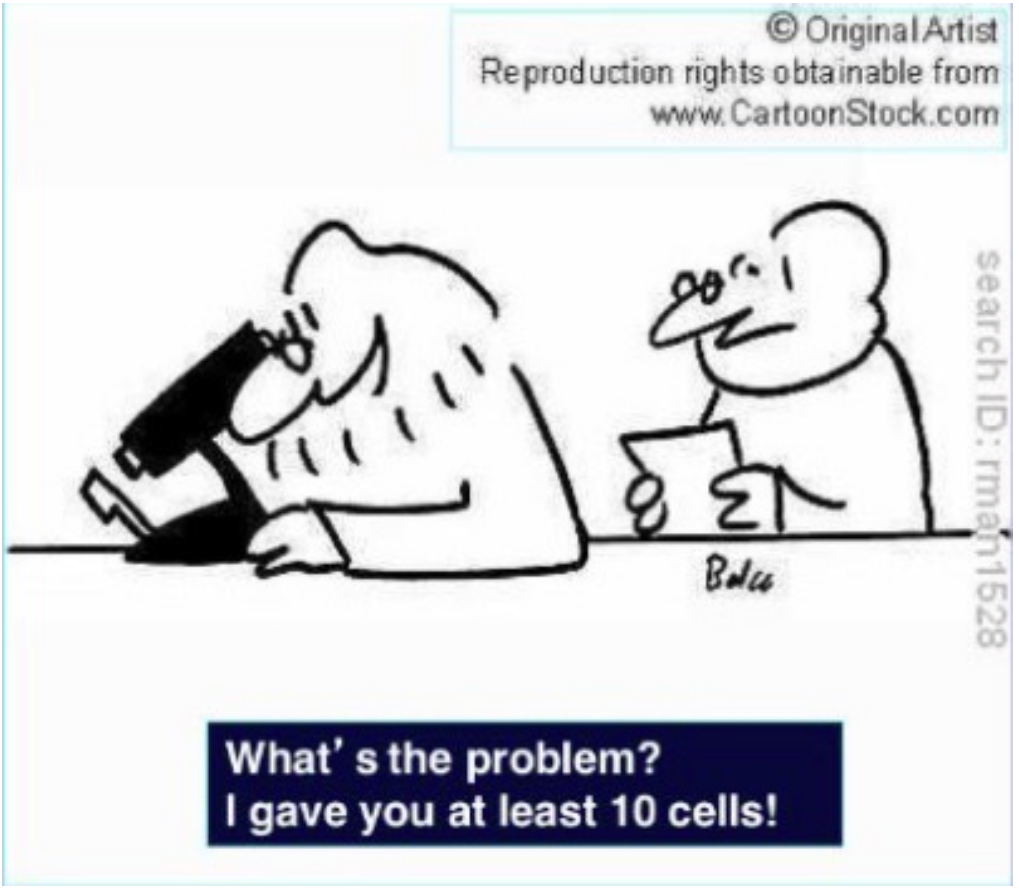


~100,000 active cancer patients across UC Health



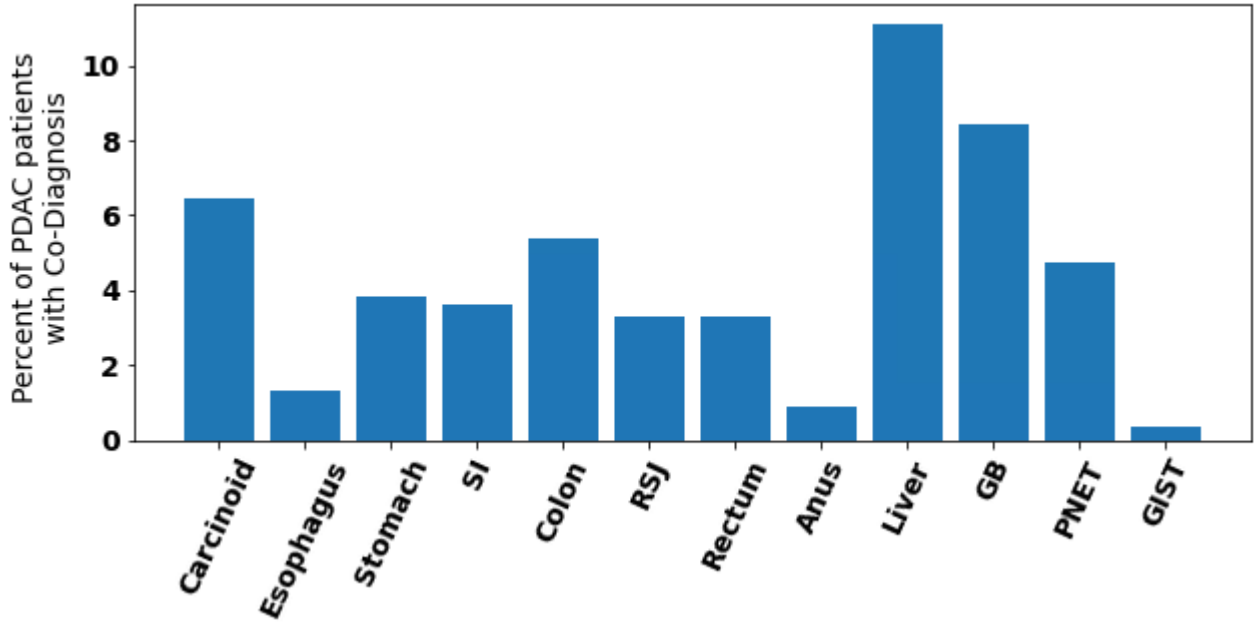
How do we identify Cancer-cohorts at scale

- Diagnosis of cancer is methodical and pedantic
  - Vast majority of Dx REQUIRE histopathological confirmation
- This is within plain text in pathology reports and challenging to process at scale across diseases
- Even if we could many patients receive Pathology outside your EMR system



Billing/Dx codes as surrogate for Diagnosis

- The most straightforward solution is to use the billing codes
- Unfortunately, this may be sensitive, but not specific
- Billing codes one of the *least* important things to a time-compressed physician
- As such, lots of noise when using Dx codes alone



Using Treatment to define cohorts in real-world Oncology can be similarly challenging

- Often involves multiple sequential lines of therapy
- Each therapy can involve multiple drugs, durations, or modalities.
- Given significant toxicities associated with treatments, therapies often have to be modified from ideal standard of care, with unclear consequences, change in day structure

Knowing the drugs doesn't identify the regimen

FLOX

FLOX: **F**luorouracil, **L**eucovorin, **OX**aliplatin

Example orders

- Example orders for FLOX in colon cancer

Regimen

Chemo

- Fluc
- Folii
- Oxa

8-week

Chemotherapy

- Fluorouracil (5-Fl
- Folinic acid (Leuc
- Oxaliplatin (Eloxa

14-day cycle for 12

FOLFOX4

FOLFOX4: **FOL**inic acid, **F**luorouracil, **OX**aliplatin

Regimen

mFOLFOX6

mFOLFOX6: **m**odified **FOL**inic acid, **F**luorouracil, **OX**aliplatin

Example orders

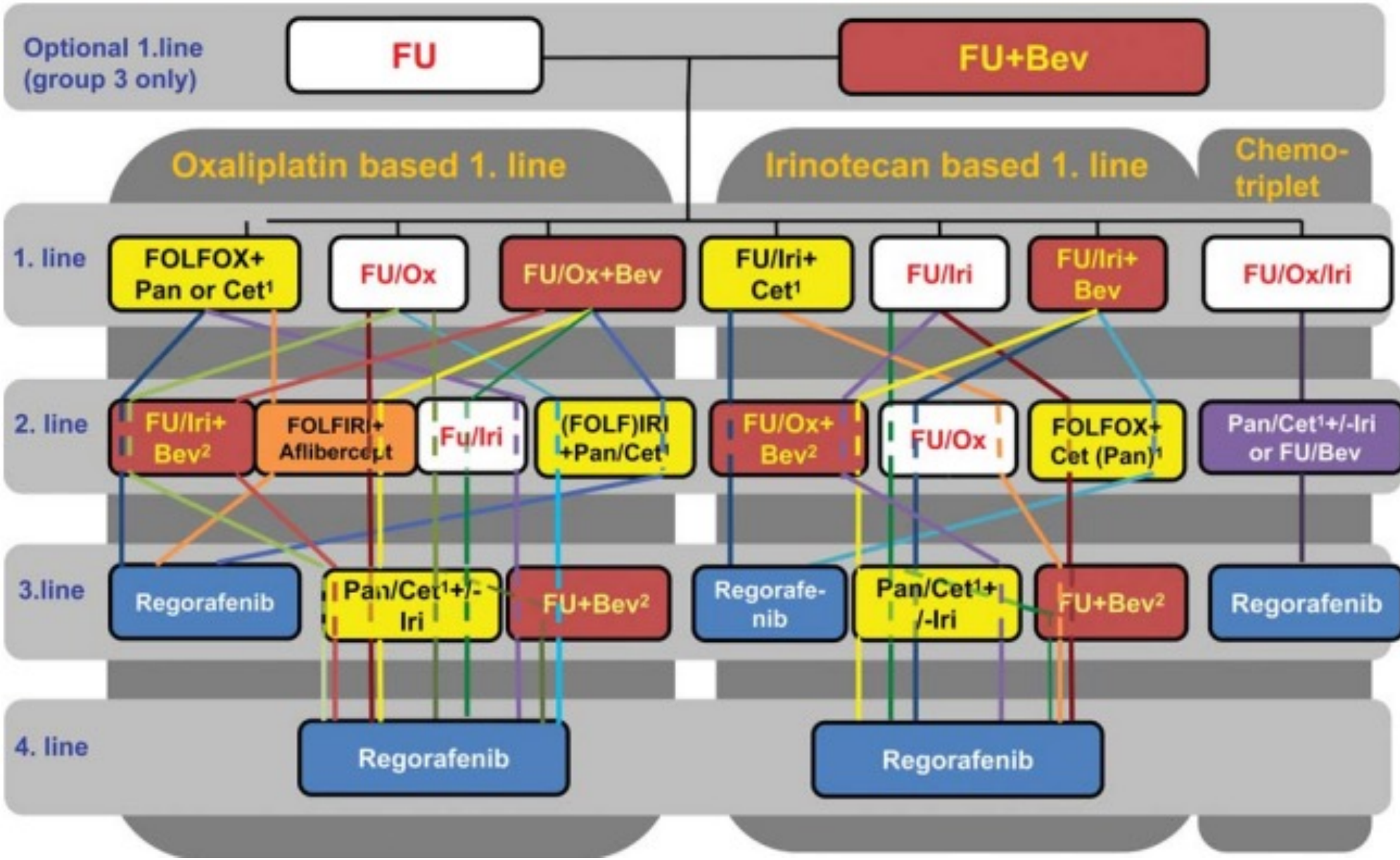
- Example orders for mFOLFOX 6 in colon cancer

Chemotherapy

- Fluorouracil (5-FU) 400 mg/m^2 IV bolus once on day 1, then 2400 mg/m^2 IV continuous infusion over 46 hours, **given second** (total dose per cycle: 2800 mg/m^2)
- Folinic acid (Leucovorin) 350 mg/m^2 IV over 2 hours once on day 1, **given first, with oxaliplatin**
- Oxaliplatin (Eloxatin) 85 mg/m^2 IV over 2 hours once on day 1, **given first, with folinic acid**

14-day cycle for 6 cycles

Cancer is an evolutionary process



Knowing the regimen doesn’t guarantee you know the treatment

- Regimens often must be modified due to side effects, toxicities, or laboratory values
- Some of these modifications involve dose, others schedule
- Sometimes whole drugs can be dropped or replaced
- Some are reversible, some are often permanent.

Leveraging Open resources for Regimen Identification

HemOnc.org - A Free Hematology/Oncology Reference

HemOnc.org is the largest freely available medical wiki of interventions, regimens, and general information relevant to the fields of hematology and oncology. It is designed for easy use and intended for healthcare professionals. Any healthcare professional can sign up to contribute; the accuracy and completeness of content is overseen by the Editorial Board. Heavily visited pages can be accessed directly from the menu on the left. If this is your first time visiting, please **go to the tutorial page** or just start exploring!

Regimens: 3,694		Regimen variants: 5,462	
Solid Tumors	Malignant Hematology	Cross-Disciplinary	Classical Hematology
Mobile Version		Desktop Version	

Links to all main disease pages

Solid Tumors			
Breast Oncology			
Breast cancer	Breast cancer, ER-positive	Breast cancer, HER2-positive	Breast cancer, ER and HER2 co-expressing
Breast cancer, triple negative (TNBC)	Breast cancer, BRCA-mutated	Breast cancer, PIK3CA-mutated	
Dermatologic Oncology			
Cutaneous BCC	Cutaneous SCC	Melanoma	Melanoma, BRAF-mutated
Melanoma, KIT-mutated	Melanoma, NRAS-mutated	Merkel cell carcinoma	Uveal melanoma

FOLFIRINOX/modified FOLFIRINOX +/- Chemoradiation

Study	Evidence
Murphy et al. 2018 (MGH 11-328)	Phase II

Note: FOLFIRINOX should be limited to those with ECOG 0-1. Amendment after first 6 patients were enrolled increased neoadjuvant cycles from 4 to 8 if no progression was detected on restaging CT

Chemotherapy

- Folinic acid (Leucovorin) 400 mg/m^2 IV over 2 hours once on day 1, **given second**
- Fluorouracil (5-FU) 400 mg/m^2 IV bolus once on day 1, then 2400 mg/m^2 IV continuous infusion over 46 hours, **given fourth** (total dose per cycle: 2800 mg/m^2)
- Irinotecan (Camptosar) 180 mg/m^2 IV over 90 minutes once on day 1, **given third** with the last 90 minutes of leucovorin; that is, irinotecan starts 30 minutes after the start of leucovorin
- Oxaliplatin (Eloxatin) 85 mg/m^2 IV over 2 hours once on day 1, **given first**

Supportive medications

- Pegfilgrastim (Neulasta) 6 mg SC once on day 4

14-day cycle for 4 to 8 cycles (see note)

FOLFIRINOX/modified FOLFIRINOX +/- Chemoradiation

Study	Evidence
Murphy et al. 2018 (MGH 11-328)	Phase II

Note: FOLFIRINOX should be limited to those with ECOG 0-1. Amendment after first 6 patients were enro

Chemotherapy

- Folinic acid (Leucovorin) 400 mg/m^2 IV over 2 hours once on day 1, **given second**
- Fluorouracil (5-FU) 400 mg/m^2 IV bolus once on day 1, then 2400 mg/m^2 IV continuous infusion over
- Irinotecan (Camptosar) 180 mg/m^2 IV over 90 minutes once on day 1, **given third** with the last 90 min
- Oxaliplatin (Eloxatin) 85 mg/m^2 IV over 2 hours once on day 1, **given first**

Supportive medications

- Pegfilgrastim (Neulasta) 6 mg SC once on day 4

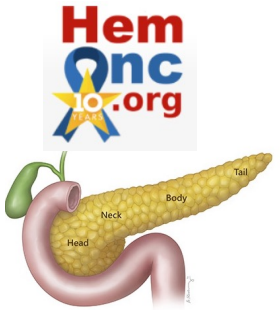
14-day cycle for 4 to 8 cycles (see note)

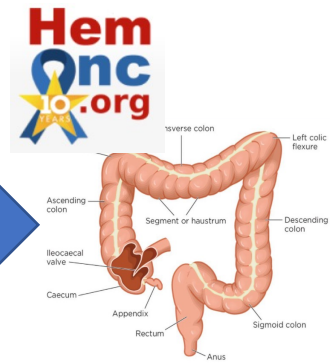
	days	fluorouracil	irinotecan	oxaliplatin
0	1	1	1	1
1	2	0	0	0
2	3	0	0	0
3	4	0	0	0
4	5	0	0	0
5	6	0	0	0
6	7	0	0	0
7	8	0	0	0
8	9	0	0	0
9	10	0	0	0
10	11	0	0	0
11	12	0	0	0
12	13	0	0	0
13	14	0	0	0

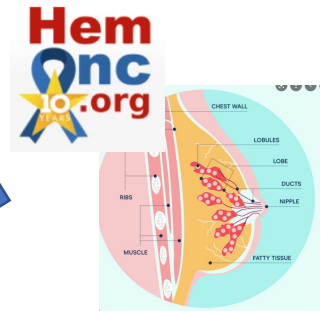
Disease-specific Regimen testing

Patient drug administration data

Drug Name
Drug A
Drug B
Drug C
Drug D
Drug E
Drug F
Drug G







Drug Name
Drug A
Drug C
Drug E

Drug Name
Drug A
Drug B
Drug C

Drug Name
Drug_G

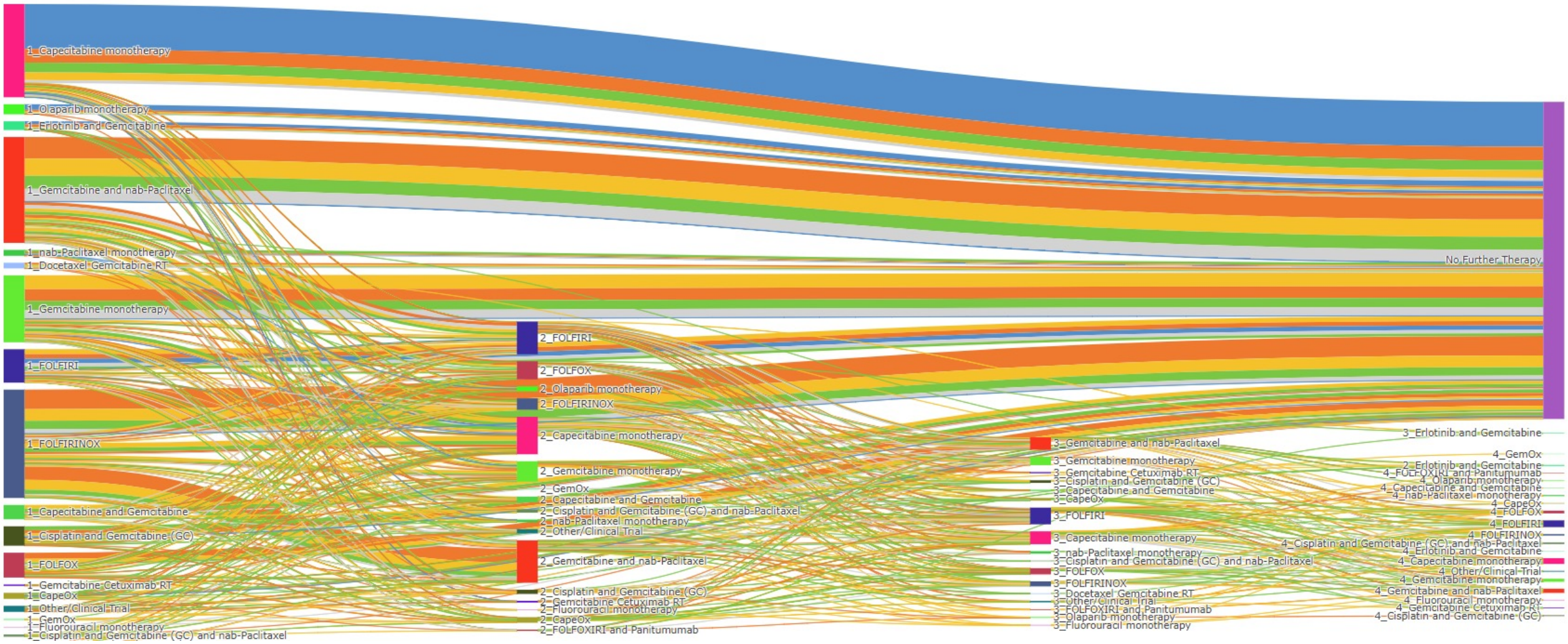
Time-Series Convolutional MLE

Overlapping, non-mutually exclusive predictions

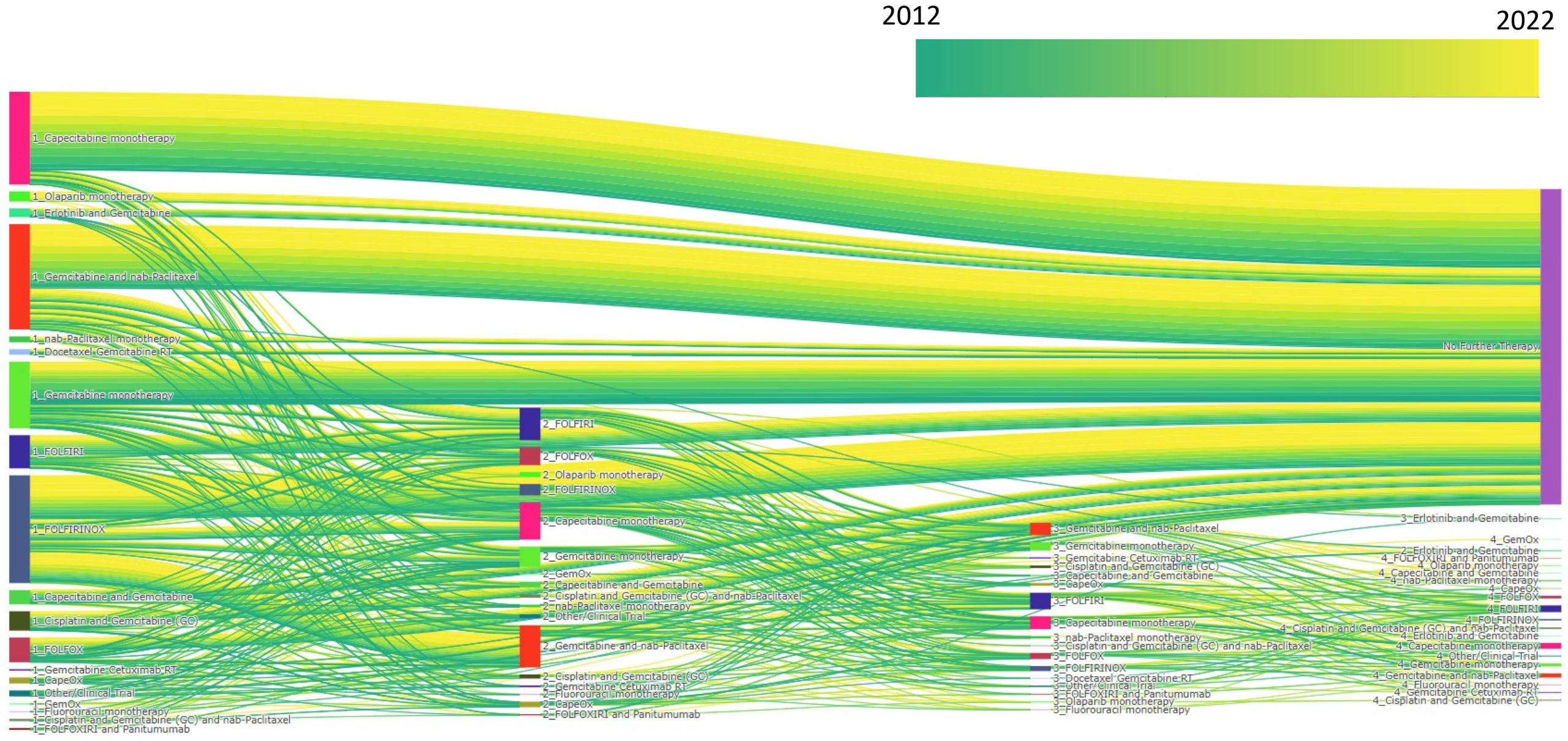
Pancreatic Treatment by Medical Center

Colored by Medical center
(5 UC medical campuses)

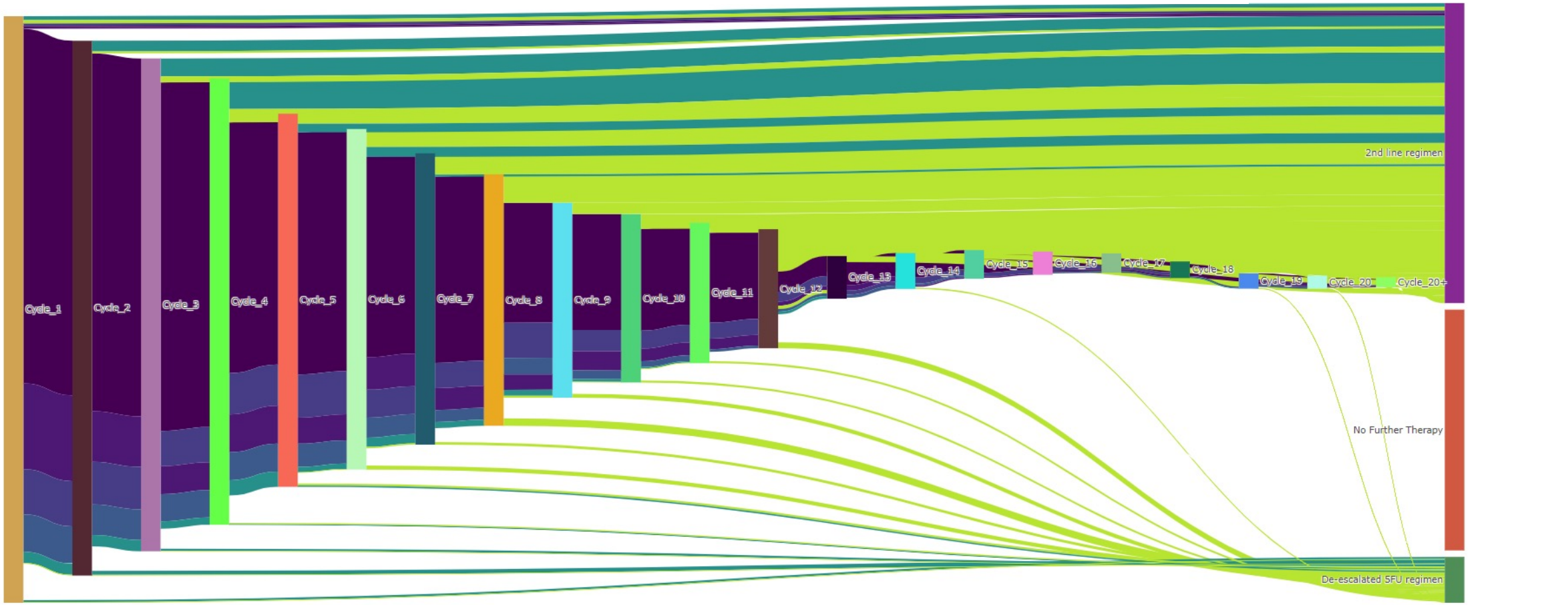
3560 PDAC patients treated with systemic therapy



Pancreatic Treatment by Year



FOLFIRINOX cycles by treatment delay



Learning patterns of treatment variation

- Gemcitabine/Abraxane is a common, highly toxic regimen

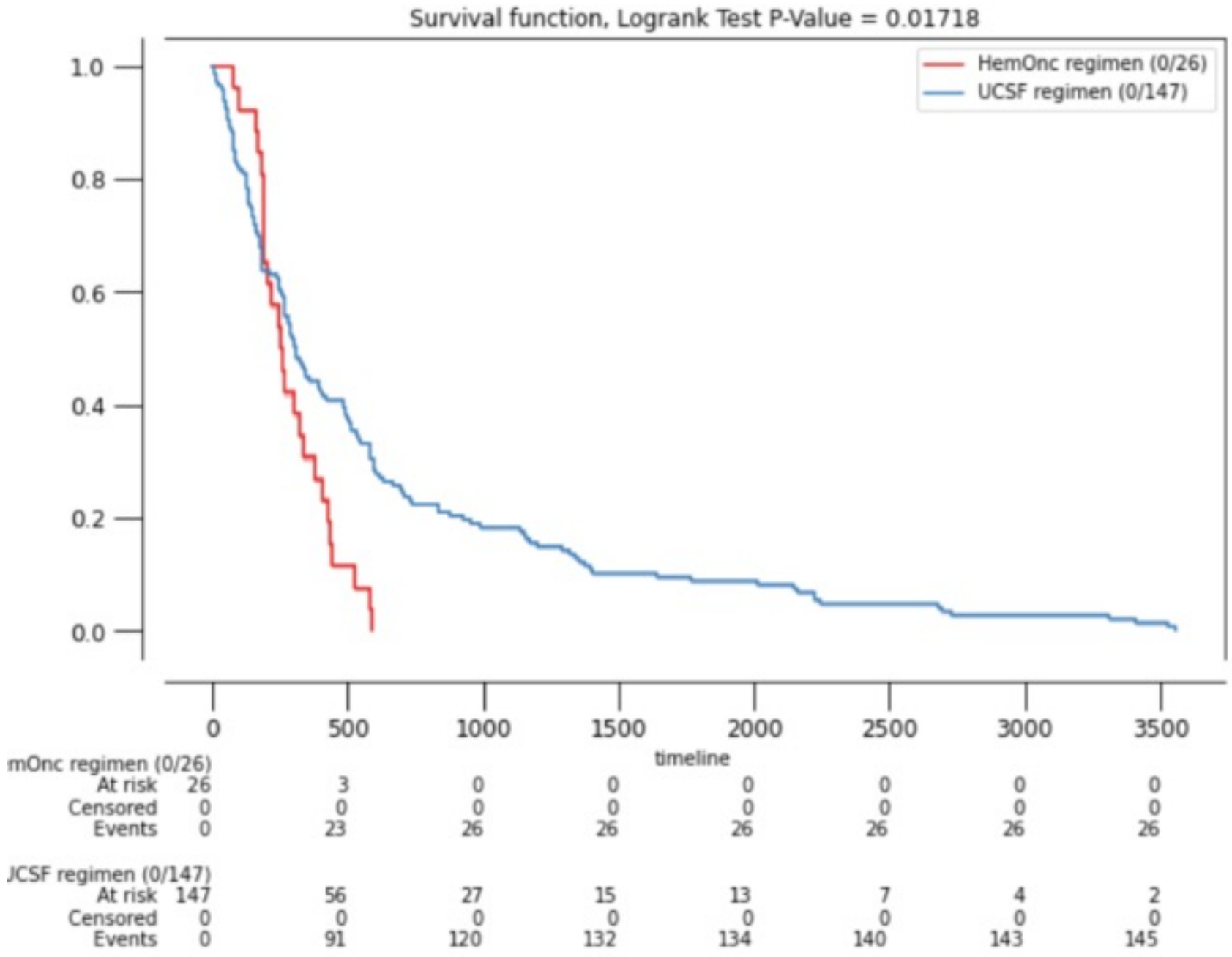
- Unsupervised learning on patients receiving this regimen identified to primary modes of treatment, one of which is not in our database.

- “Off-label” regimen efficacy

	gemcitabine	abraxane
0	1	1
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	1	1
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	1	1
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0
21	0	0
22	0	0
23	0	0
24	0	0
25	0	0
26	0	0
27	0	0

	date	gemcitabine	abraxane
43	2018-12-16	1	1
44	2018-12-17	0	0
45	2018-12-18	0	0
46	2018-12-19	0	0
47	2018-12-20	0	0
48	2018-12-21	0	0
49	2018-12-22	0	0
50	2018-12-23	0	0
51	2018-12-24	0	0
52	2018-12-25	0	0
53	2018-12-26	0	0
54	2018-12-27	0	0
55	2018-12-28	0	0
56	2018-12-29	0	0
57	2018-12-30	1	1
58	2018-12-31	0	0
59	2019-01-01	0	0
60	2019-01-02	0	0
61	2019-01-03	0	0
62	2019-01-04	0	0
63	2019-01-05	0	0
64	2019-01-06	0	0
65	2019-01-07	0	0
66	2019-01-08	0	0
67	2019-01-09	0	0
68	2019-01-10	0	0
69	2019-01-11	0	0
70	2019-01-12	0	0

Are we giving patients too much Chemotherapy?: Learning from off-label regimen use



Next Steps

- Automated unsupervised learning across disease and detected regimens to identify “modes” of treatment modification

- Using variability in Real-world clinical practice as natural experimentation to identify Regimens where de-escalation may NOT affect efficacy.

- Create pan-cancer clinical-decision-support models to predict toxicity and intolerance in traditional chemotherapy utilization.

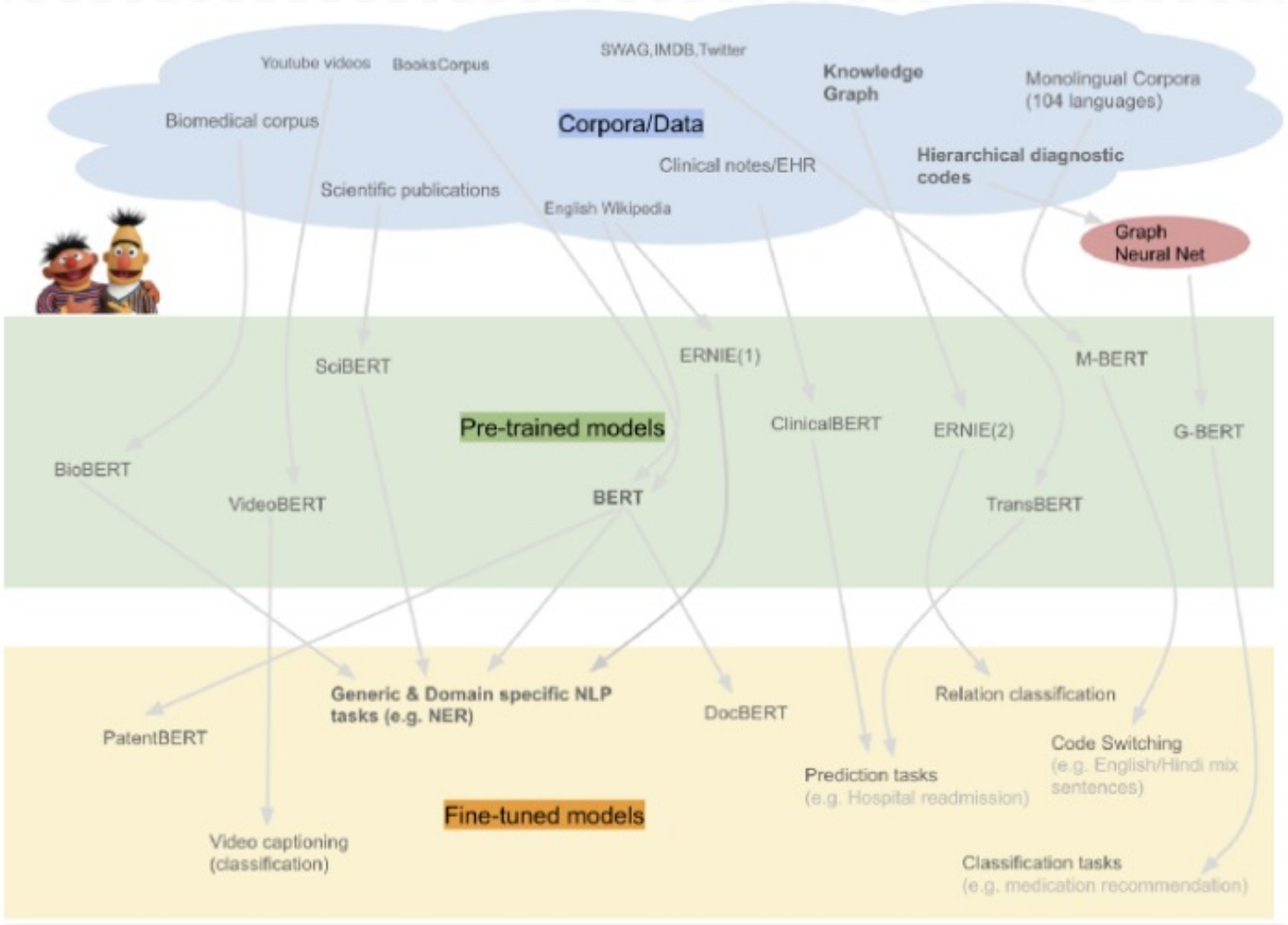
Accurate extraction of rich oncology diagnosis, treatment and response data from oncology notes

Oncologic History: Physicians attempt to summarize all data relevant to patient oncology course

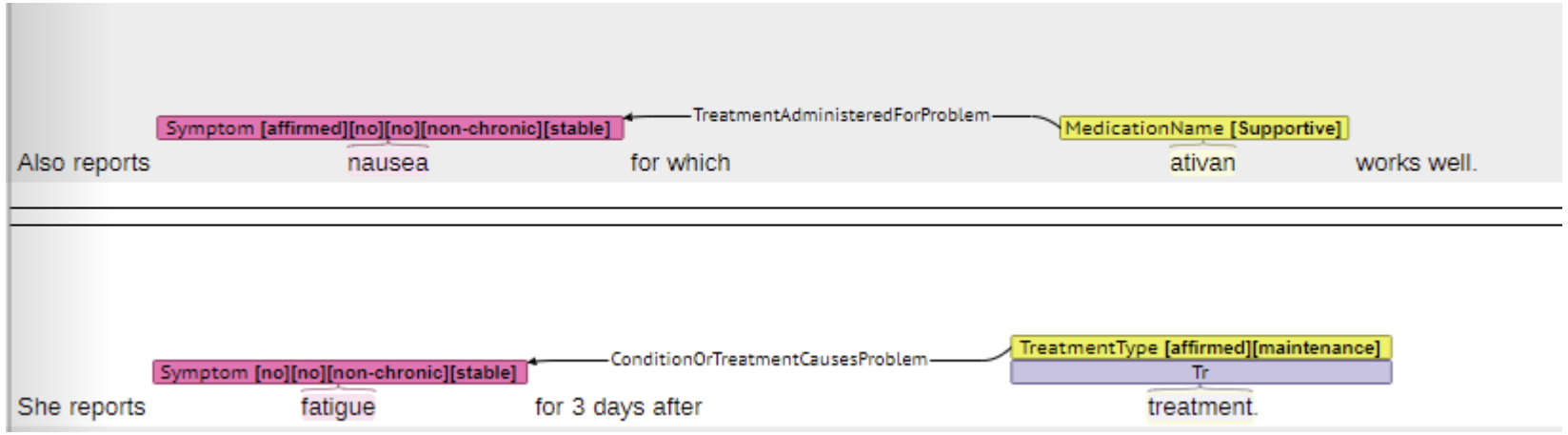
- Reason patient originally presented
- How diagnoses was made and initial staging
- Previous treatments and responses
- Patient toxicity and trajectory

Recently, huge leaps in AI approaches to NLP

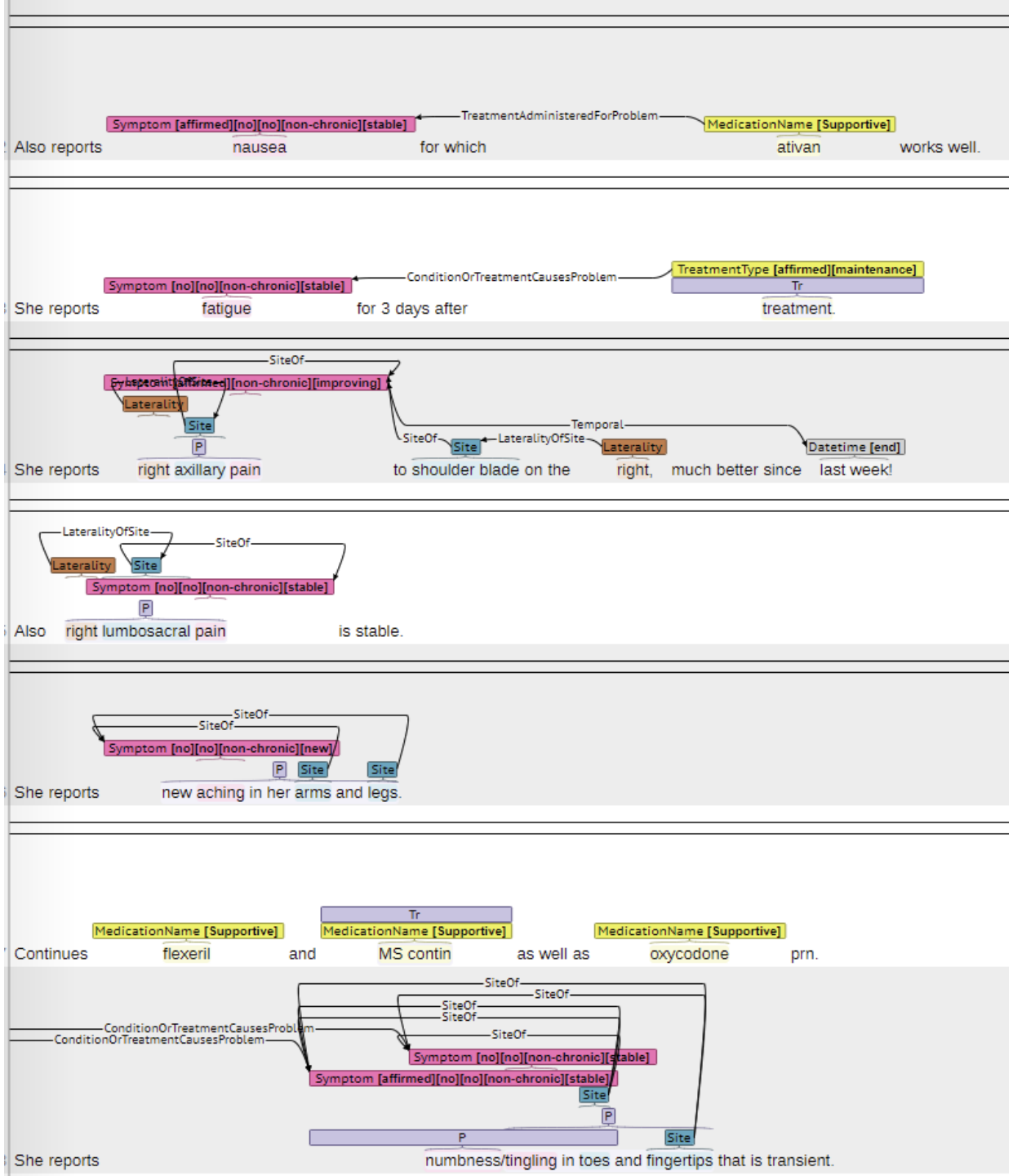
Strong performance in ability to understand complex relations in grammar and language



Identifying causality and “physician-belief”



Identifying causality and “physician-belief”



Acknowledgements

Butte Lab

- **Prof. Atul Butte**
- **Debajyoti Datta**
- Ben Rubin
- Rohit Vashisht
- Ayan Patel
- Doug Arneson
- **Vivek Rudrapatna**
- Zicheng Hu
- Sanchita Bhattacharya
- Boris Oskotsky
- Gregory Goldgof
- Andrew Bishara
- **Madhumita Sushil**
- Kendra Radtke
- Daniel Wong
- Charlotte Nelson
- Marie Binvignant
- Brenda Miao
- Harry Sun
- **Divneet Mandair**

Gundolf Schenk

Sharat Israni

Prof. Eric Collisson

Prof. Julian Hong

Prof. Margaret Tempero

Prof. Atul Butte

UCSF Bakar Computational Health Sciences Institute